# Phylogenetic analysis of HIV-1 in Mpumalanga

Wela Patrick Msimanga

Thesis presented in fulfillment of the requirements for the degree Masters of Sciences in Medical Sciences (Medical Virology) at the Faculty of Medicine and Health Sciences, Stellenbosch University


Division of Medical Virology, Department of Pathology

Supervisor: Professor Susan Engelbrecht

March 2013

# Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Wela Patrick Msimanga
**Name in full**

___ 06 __/__ 12 __/__ 2012 _
**Date**

# Acknowledgements

I would hereby like to express my full gratitude to the following people and institutions; without whom this thesis would not have been possible:

- Prof. Susan Engelbrecht, the supervisor of this study, for her expert advice, guidance and assistance throughout the course of this project.
- To my mother, Evelyn Hesmie Msimanga, for her support and forbearance.
- To all the staff and students of the Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, for their moral support during the course of this study.
- To all the staff of the National Institute for Communicable Diseases, Sandringham, Johannesburg, for their moral support throughout the years.
- To the Poliomyelitis Research Foundation (PRF), Department of Health and European Union for their generous financial support.

To the Poliomyelitis Research Foundation, as well as the Faculty of Medicine and Health Sciences of Stellenbosch University, for the grant money that helped in funding this study

# Abstract

The diversity of HIV-1 sequences derived from patients in Bushbuckridge, Mpumalanga, was investigated. The *gag* p24, *pol* p10 and p66/p51, *pol* p31 and *env* gp41 gene fragments from 51 patients were amplified and sequenced. Quality control on the sequences was carried out using the LANL QC online tool. HIV-1 subtype was assigned using the LANL QC (RIP), REGA and jpHMM online tools. Subtype for the *pol* gene fragment was further designated using the SCUEAL online tool. Most of the sequences, that is 89%, belonged to HIV-1 subtype C. LANL QC (RIP), REGA, jpHMM also detected recombinants in 11% of the sequences. One of the isolates could only have the *env* gp41 gene fragment amplified and sequenced, which was determined to be HIV-1 subtype B. Phylogenetic analysis using the Neighbor-Joining and Maximum Likelihood methods from MEGA v 5 showed that, except for the *env* gp41 designated as a subtype B, all sequences in the study clustered with HIV-1 subtype C. Significantly, phylogenetic analysis showed that not only are the Bushbuckridge, Mpumalanga sequences related to HIV-1 subtype C sequences from southern Africa, India, Ethiopia and Brazil, but it is possible there has been multiple introductions of HIV-1 in the province. SDRMs were observed in two samples.

# Presentations

Engelbrecht S, Msimanga P, Wilkinson E, Glashoff R, Vardas E.
**HIV-1 recombinants in Mpumalanga Province, South Africa.**
17th International HIV Dynamics and Evolution Conference, Asilomar Conference Grounds, Pacific Grove, CA, USA, 25 – 28 April, 2010.

Engelbrecht S, Isaacs S, Msimanga P, Vardas E, Glashoff R.
**HIV-1 recombinants identified in Mpumalanga and Eastern Cape Provinces of South Africa.**
10th International Conference on Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases (MEEGID X), Amsterdam, the Netherlands, 3 – 5 November 2010.

# TABLE OF CONTENTS

## List of Abbreviations

| | |
|---|---|
| A | Adenine |
| AIDS | Acquired Immune Deficiency Syndrome |
| AMV | Avian Myeloblastosis Virus |
| ARV | Antiretroviral |
| ART | Antiretroviral therapy |
| ATP | Adenosine triphosphate |
| AZT | Azidothymidine |
| BI | Bayesian Inference |
| BLAST | Basic Local Alignment Search Tool |
| bp | Base pairs |
| BR | Brazil |
| BW | Botswana |
| C | Cytosine |
| CA | Capsid protein p24 |
| CCR5 | C-C chemokine receptor 5 (β-chemokines) |
| CD | Democratic Republic of Congo |
| CD4 | Cluster of differentiation 4 |
| CDC | Centers for Disease Control and Prevention |
| CDK9 | cyclin-dependent kinase 9 |
| cDNA | complimentary DNA |
| CF | Central African Republic |
| CM | Cameroon |
| CPR | Calibrated Population Resistance |
| CRF | Circulating Recombinant Form |
| CTD | C-terminal domain |
| CXCR4 | C-X-C chemokine receptor 4 (α-chemokines) |
| D | Aspartic acid |
| DIS | Dimer initiation signal |
| DNA | Deoxyribonucleic acid |
| dNTP | deoxynucleoside triphosphate |
| DRM | drug resistance mutations |
| EDTA | Ethylenediaminetetraacetic acid |

| | |
|---|---|
| EFV | Efavirenz |
| *env* | Envelope gene |
| ET | Ethiopia |
| *et al* | *et alibi* |
| FEL | Felsenstein (1981) model of evolution |
| FDA | Food and Drug Administration |
| FR | France |
| FM | Fitch-Margoliash's least squares method |
| G | Guanine |
| *gag* | Group specific antigen gene |
| GTR | General time-reversible (1986) model of evolution |
| HAART | Highly active antiretroviral therapy |
| HIV-1 | Human immunodeficiency virus type 1 |
| HIV-2 | Human immunodeficiency virus type 2 |
| HKY | Hasegawa-Kishino-Yano (1985) model of evolution |
| HLA | Human leukocyte antigen |
| HTLV-I | Human T-cell leukemia virus I |
| HTLV-II | Human T-cell leukemia virus II |
| ICAM-1 | Intercellular Adhesion Molecule-1 |
| ICTV | International Committee on the Taxonomy of Viruses |
| IDU | Injecting drug use |
| IN | India |
| IN | Integrase p32 |
| JC | Jukes and Cantor (1969) model of evolution |
| jpHMM | jumping profile Hidden Markov Model |
| K2P | Kimura 2 parameter (1980) model of evolution |
| KE | Kenya |
| LANL | Los Alamos National Laboratory |
| LAV | Lymphadenopathy-associated virus |
| LCA | Last Common Ancestor |
| LTR | Long terminal repeat |
| MA | Matrix protein p17 |
| MCMC | Markov chain Monte Carlo |
| ME | Minimum Evolution |

| | |
|---|---|
| $MgCl_2$ | Magnesium Chloride |
| MHC I | Major Histocompatibility Class I |
| MHC II | Major Histocompatibility Class II |
| MI | Michigan |
| ML | Maximum Likelihood |
| MMWR | Morbidity and Mortality Weekly Report |
| MP | Maximum Parsimony |
| MRCA | Most Recent Common Ancestor |
| mRNA | messenger Ribonucleic acid |
| M-tropic | Macrophage tropic |
| MW | Malawi |
| NASBA | Nucleic acid sequence-based amplification |
| NAM | nucleoside analogue-associated mutation |
| NC | Nucleocapsid protein p7 |
| *nef* | Negative effector gene |
| NF-κB | Nuclear factor kappa B |
| NG | Nigeria |
| NJ | Neighbor Joining |
| NNRTI | Non-nucleoside reverse transcriptase inhibitor |
| NRTI | Nucleoside reverse transcriptase inhibitor |
| NSI | Non-syncytium inducing |
| $OH^-$ | Hydroxyl group |
| ORF | Open Reading Frame |
| PBL | Peripheral blood lymphcytes |
| PBMC | Peripheral blood mononuclear cells |
| PBS | Primer binding site |
| PCR | Polymerase Chain Reaction |
| PI | Protease Inhibitor |
| PIC | Pre-integration complex |
| PIME | Prevalence, Incidence and Molecular Epidemiology Study |
| *pol* | Polymerase gene |
| PPT | Polypurine tract |
| PR | Protease p10 |
| PRT | Phenotypic Resistance Test |

| | |
|---|---|
| R | Repeat |
| RIP | Recombination Identification Programme |
| *rev* | Regulator of expression of viral proteins gene |
| RNase H | Ribonuclease H |
| RNA | Ribonucleic acid |
| RNA Pol II | Ribonucleic acid Polymerase II |
| RNP | Ribonucleoprotein |
| RRE | Rev-response element |
| RT | Reverse transcriptase p66/p51 |
| RT-PCR | Reverse Transcriptase Polymerase Chain Reaction |
| RTC | Reverse transcription complex |
| SDMR | Surveillance drug resistance mutations |
| SE | Sweden |
| SGR | Second Generation Recombinants |
| SI | Syncytium-inducing |
| SIV | Simian immunodeficiency virus |
| SIVcpz | Simian immunodeficiency virus chimpanzee |
| SIVsm | Simian immunodeficiency virus sooty mangabey |
| ssDNA | single-stranded Deoxyribonucleic acid |
| STI | Sexually transmitted infection |
| SU | Surface glycoprotein gp120 |
| SYM | Symmetrical (1994) model of evolution |
| T | Thymine |
| TAMs | Thymidine analogue-associated mutations |
| TAR | Transactivation response region |
| T92 | Tamura (1992) model of evolution |
| T-cell | Thymus-derived lymphocyte |
| TAR | Transactivation response region |
| *Taq* | *Thermus aquaticus* |
| *tat* | Trans-activator of transcription gene |
| TDF | Tenofovir disoproxil fumarate |
| TE | Tris EDTA |
| *Tfl* | *Thermus flavus* |
| TH | Thailand |

| | |
|---|---|
| TM | Transmembrane glycoprotein gp41 |
| TN93 | Tamura and Nei (1993) model of evolution |
| T-tropic | T-lymphocyte tropic |
| Tris | Tris(hydroxymethyl)aminomethane |
| $tRNA_{lys3}$ | lysine transfer RNA |
| TZ | Tanzania |
| U | Uracil |
| UNAIDS | Joint United Nations Programme on HIV/AIDS |
| UNDP | United Nations Development Programme |
| UNGASS | United Nations General Assembly Special Session |
| U3 | Unique 3′ |
| U5 | Unique 5′ |
| UPGMA | Unweighted Pair Group Method with Arithmetic Mean |
| URF | Unique Recombinant Form |
| US | United States of America |
| UV | Ultra Violet |
| V | Volts |
| VL | Viral load |
| *vif* | Virion infectivity factor gene |
| *vpr* | Viral protein, regulatory gene |
| *vpu* | Viral protein, unknown gene |
| WHO | World Health Organization |
| WI | Wisconsin |
| ZA | South Africa |
| ZM | Zambia |

**List of Symbols**

| | |
|---|---|
| α | rate of transition substitutions |
| α | shape parameter for gamma-distributed rates among sites |
| β | scale parameter for gamma-distributed rates among sites |
| β | rate of transversion substitutions |
| © | Copyright |
| cm | centimeter |
| °C | degree Celsius |

| | |
|---|---|
| Γ | gamma-distributed rate variation among sites |
| g | gram |
| *g* | unit of gravity |
| κ | ratio of transition/transversion rates |
| kb | kilobases |
| kD | kiloDaltons |
| M | Molar |
| m | meter |
| µg | microgram |
| µl | microliter |
| ml | milliliter |
| mM | milliMolar |
| µM | microMolar |
| µm | micrometer |
| mol | mole |
| ng | nanogram |
| nm | nanometer |
| $\pi_A$ | base frequency of adenine |
| $\pi_C$ | base frequency of cytosine |
| $\pi_G$ | base frequency of guanine |
| $\pi_T$ | base frequency of thymidine |
| pmol | Picomole |
| ® | Registered |
| ™ | Trade Mark |
| $T_A$ | Annealing temperature |
| $T_M$ | Melting temperature |
| U | unit |
| τ | tree topology |
| θ | parameters of substitution model |
| χ | aligned sequences |
| ν | set of branch lengths |

## List of Figures

## List of Tables

# CHAPTER ONE: INTRODUCTION AND LITERATURE REVIEW

# CHAPTER ONE

# INTRODUCTION AND LITERATURE REVIEW

## INTRODUCTION

Chapter one begins with an overview of the devastating public health impact of the HIV/AIDS pandemic. The subsequent sections seek to review the literature on aspects of HIV-1 relevant to the focus of the study, which is the phylogenetic analysis of HIV in Bushbuckridge, Mpumalanga. These aspects include the origins of HIV/AIDS; the biology of HIV including its taxonomy, morphology, genomic structure and replication, and the nature, causes and public health implications of HIV diversity. The literature review concludes with an examination of the molecular epidemiology of HIV both in South Africa and globally. The final section of Chapter 1 considers concepts and tools in HIV-1 phylogenetic analysis.

### The HIV/AIDS pandemic

By 2011 there were an estimated 34.2 million million people living with HIV. An estimated 2.5 million people had become infected with the virus and 1.7 million people had died of HIV-related causes in 2011 alone (UNAIDS (2012) Report on the global Aids epidemic Geneva UNAIDS). Twenty five million people have died of HIV-related causes since the beginning of the epidemic. While Sub-Saharan Africa accounts for 11–12% of the world's population, it is home to 68.7% of people living with HIV (**Figure 1.1**) (UNAIDS, 2012. Report on the global Aids epidemic Geneva UNAIDS).

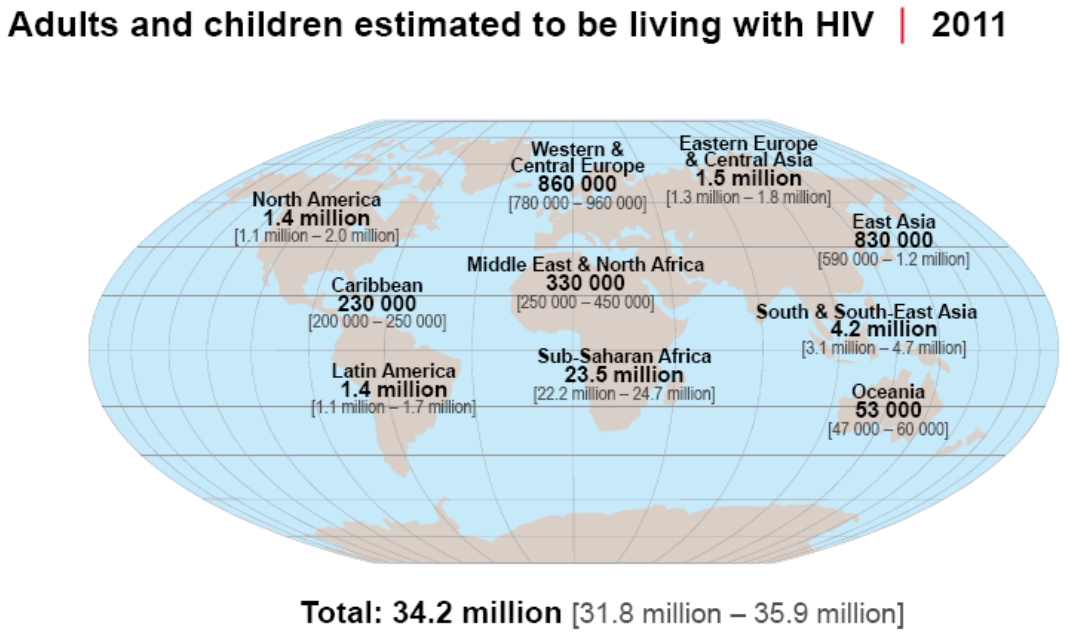



**Figure 1.1.** Adults and children estimated to be living with HIV in 2011. (Source: UNAIDS, 2012).

### The HIV/Aids epidemic in South Africa

South Africa, with an estimated 5.63 million South Africans living with HIV in 2009, has the highest number of people living with HIV in the world, and almost one in five South African adults is HIV-positive. There has been a substantial decrease in AIDS related mortality with the annual number of AIDS deaths reduced from about 257,000 in 2005 to about 194,000 in 2010. Nationally, the HIV prevalence among women in the age group 30 - 34 years remains the highest with a prevalence of 42.6% in 2010. In 2008, HIV prevalence among women aged 20-24 was 21.1% compared to 5.15% in men aged 20-24 years. The prevalence of HIV in South Africa shows considerable variance across its nine provinces, with the highest HIV prevalence among the 15-49 year olds in KwaZulu-Natal at 39.5% and the lowest Northern Cape at 18.4% (South African Government, 2012).

### Bushbuckridge, Mpumalanga

The Bushbuckridge Local Municipality is a predominantly rural area with a few small urban centres that include, Thulamahashe, Acornhoek, Marite, Dwarsloop and Mkhuhlu. The Bushbuckridge Local Municipality is situated in

the Ehlanzeni District Municipality, which is in the north east of Mpumalanga Province, an area of the lowveld bordered by the Kruger National Park in the east, the Limpopo Province in the north and the Blyde River Canyon in the west. The population of Bushbuckridge is very poor, with only 14% of the adult population employed and over 85% of households living below the household subsistence level (HSL). Half of males and 14% of females between the ages of 25 and 59 are long-term migrant workers and represent a source of remittances, which comprise the largest proportion of the income of the population of Bushbuckridge (South African Government).

The overall HIV prevalence in Mpumalanga in 2010 was 35.1%, slightly less than the 37.7% prevalence in Ehlanzeni District. The Gert Sibande District had the highest prevalence at 38.8%, and the Nkangala District had the lowest at 27.7% (South African Government, 2010).

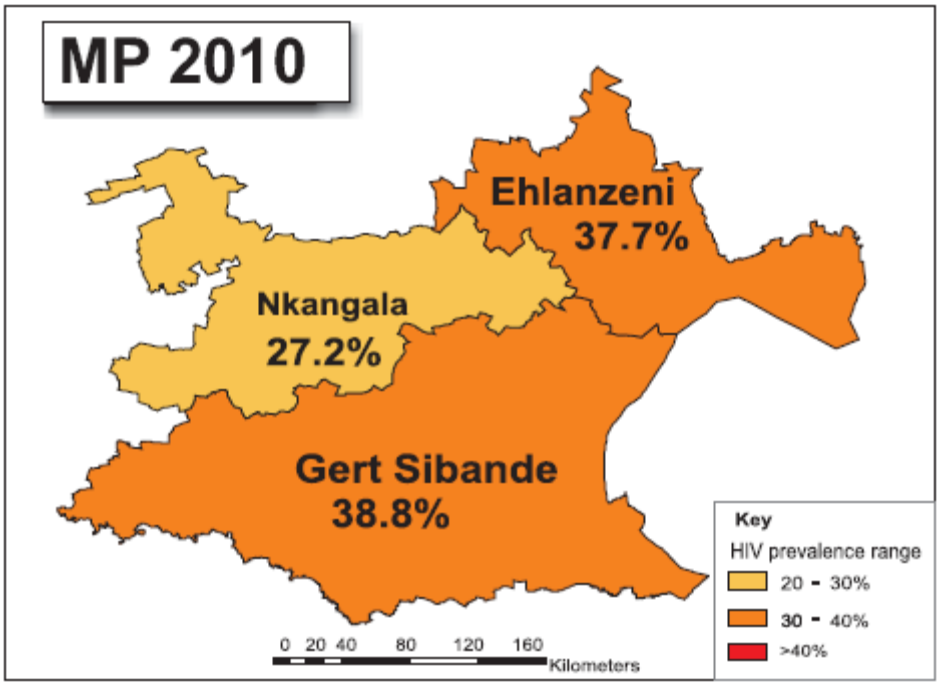



**Figure 1.2.** HIV prevalence distribution among antenatal women by district, Mpumalanga, 2010 (Source: South African Government, 2010).

HIV prevalence in South Africa is characterized by extreme heterogeneity, both between and within provinces. There is, for example, highly uneven HIV prevalence between districts in the Western Cape Province, with districts comprising predominantly informal areas having the highest overall HIV

prevalence (South African Government, 2007). There is also considerable variation in HIV prevalence amongst South Africa's provinces, with KwaZulu-Natal having the highest antenatal prevalence in 2005 at 39.1% and the Western Cape the lowest at 15.7% (South African Government, 2010).

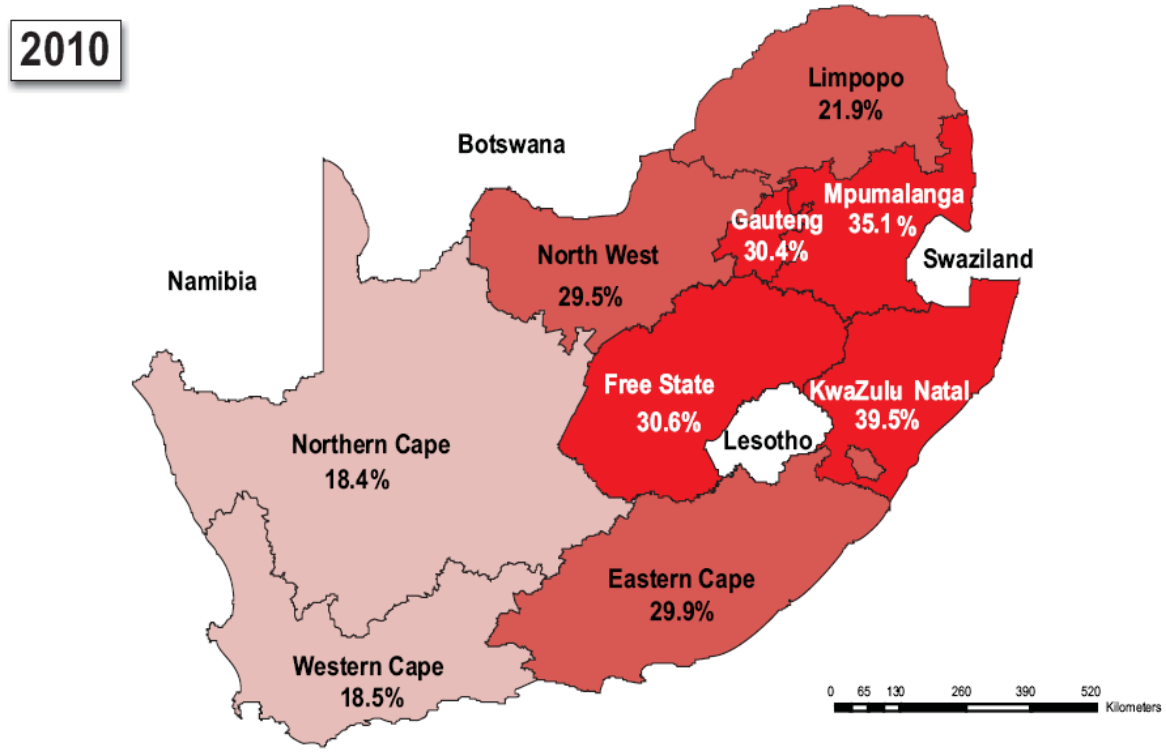



**Figure 1.3.** HIV prevalence distribution by province, South Africa 2010 (Source: South African Government, 2010).

While the overall estimate of HIV prevalence in the general population in South Africa for 2010 was 17.9, the national HIV prevalence estimate among antenatal women in 2010 was 30.2% (South African Government, 2010).

## LITERATURE REVIEW

### 1. The origins and history of HIV/AIDS

In 1981 the Morbidity and Mortality Weekly Report (MMWR) of the Centers for Disease Control and Prevention (CDC) in the United States (US) reported cases of *Pneumocystis carinii* pneumonia, Kaposi's sarcoma and T-lymphocyte dysfunction in previously healthy young gay men and intravenous drug users (IDUs) in California and New York. When the CDC eventually

defined the Acquired Immnune Deficiency Syndrome (AIDS) by 1982, a similar immunodeficiency had been recognized in persons with haemophilia A and heterosexual Haitians who had immigrated to the United States. The AIDS epidemic in Africa was not recognized until 1984. While the AIDS epidemiv in Europe was similar to the one in the Uited States in so far as it predominantly affected homosexual males, the epidemic in Africa replicated the situation with Haitian immigrants in the US and impacted mostly heterosexuals (CDC, 1981; CDC, 1982a; CDC, 1982b; Apetrei et al., 2004).

In 1983 Barre-Sinoussi and her colleagues (1983) had isolated a retrovirus from the lymph node cells of a patient with lymphadenopathy and, accordingly, the virus was designated lymphadenopathy-associated virus (LAV). In 1986, a novel retrovirus similar to HIV, but also clearly different, was isolated from an immunocompromised individual in West Africa. The new retrovirus was initially called lymphadenopathy-associated virus-2, but its name was subsequently changed to HIV-2. this entailed that the aetiological agent of AIDS in the United States, Haiti, Europe, and central Africa should be renamed human immunodeficiency virus type 1 (HIV-1). HIV-2 is endemic in West Africa and has not spread pandemically, unlike HIV-1. HIV-2 is transmitted at lower rates than HIV-1. Further, HIV-2 infection produces lower plasma viral loads, causes AIDS in a lower percentage of patients, and has a longer incubation period (Apetrei et al., 2004).

## 2. The HIV-1 virion and life cycle

### 2.1 HIV taxonomy

According to the classification system of the International Committee on the Taxonomy of Viruses (ICTV), HIV belongs to the genus *Lentivirus*, subfamily orthoretrovirinae, and family *Retroviridae*. The designation lentivirus is derived from the Latin word *lentus,* which means ‘slow,’ since *lentiviruses* differ from other retroviruses by their long latent period before the manifestation of clinical illness. Lentiviruses also differ from other retroviruses by their complicated genome structures (Wang et al., 2000). Classification of *Retroviridae* is based on the phylogenetic analysis of the most conserved

regions of reverse transcriptase (Coffin, 1999). *Retroviridae* is a family of single-stranded RNA viruses that utilize reverse transcriptase to synthesize DNA by using RNA as a template, which is then integrated into the genome of infected cells. *Retroviridae* are also diploid in that they possess two genomic RNA molecules per virion (Jetzt et al., 2000). According to the Baltimore system for the taxonomic classification of viruses HIV, as a single-stranded positive sense RNA virus with a DNA intermediate in its life-cycle, belongs to Group VI.

Lentiviruses are characterized by the morphogenesis of the virion at the cell surface, the assembly of core components concomitantly with the budding process, and an elongated core comprising a core shell and the RNP nucleoid (Gelderblom et al., 1987). Lentiviruses also possess the ability to productively infect some types of non-dividing cells (Freed, 2001). By virtue of the fact that first, they utilize viral factors to regulate their expression, and second, they encode regulatory and accessory proteins essential to their life cycle, lentiviruses are described as complex retroviruses (Kuiken, et al., 2008).

### 2.2 HIV morphology

HIV-1 virions are spherical to pleomorphic in shape, 80 nm to 100 nm in diameter and possess a lipid bilayer envelope derived from the membrane of the host cell (**Figure 1.4**). The envelope contains 72 projections comprising the surface (SU) glycoprotein that is anchored to the lipid bilayer through interactions with the transmembrane (TM) glycoprotein. The lipid bilayer also contains several cellular membrane proteins derived from the host cell, including the major histocompatibility complex I and II (MHCI and II) proteins, and the intercellular adhesion molecule-1 (ICAM-1).

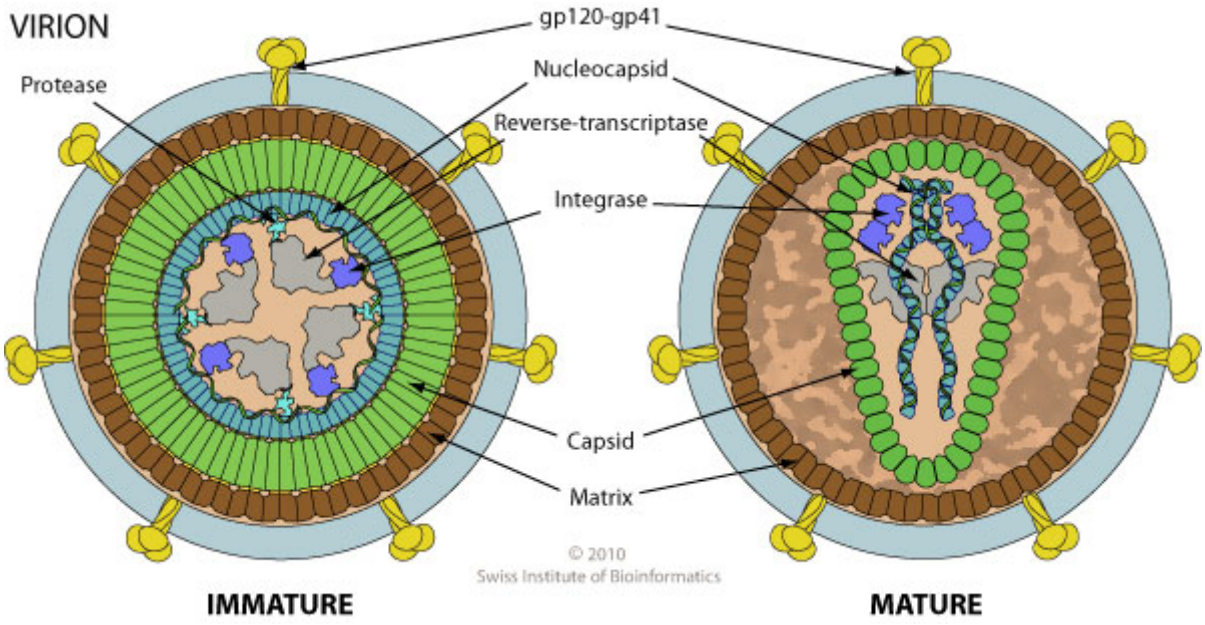



**Figure 1.4.** Structures of immature and mature HIV virions (Source: http://www.expasy.org/viralzone, Swiss Institute of Bioinformatics).

A layer comprising the matrix protein MA is embedded in the inner surface of the lipid bilayer, and a conical-shaped capsid core comprising the capsid (CA) is located in the center of the virus. The capsid encloses two copies of the unspliced viral genome, which in turn is stabilized as a ribonucleoprotein complex through interaction with the NC. The capsid also contains protease (PR), reverse transcriptase (RT) and integrase (IN). Virus particles further enclose the accessory proteins, Nef, Vif and Vpr. Three additional accessory proteins that function in the host cell, Rev, Tat and Vpu, do not appear to be contained in the capsid (Briggs et al., 2003; Rubbert et al., 2007; Turner and Summers, 1999).

### 2.3 HIV genome structure

The HIV-1 genome (**Figure 1.5**) comprises a homodimer of linear, positive-sense, single-stranded RNA that is 9.181 kb in size (Zhuang et al., 2002). The two RNA strands are capped, polyadenylated and noncovalently joined at a sequence near the 5' end of the genome known as the dimer initiation signal

(DIS). Lysine transfer RNA ($tRNA_{lys3}$) is bound to the viral RNA in the viral particle and serves as a primer for reverse transcription.

The basic, common structure of the provirus of primate lentiviruses is LTR-*gag-pol-vif-vpr-tat-rev-env-nef*-LTR (Beer, Bailes, Sharp, Hirsch, 1999). The long terminal repeats (LTRs) contain elements that are involved in the regulation of the transcription of the provirus, including promoter sequences, and they comprise U3, R and U5 (unique 3′, repeat, unique 5′ sequence). The generation of the viral genomic RNA results in the loss of the promoter sequences at their 5′ ends (Hu and Temin, 1990; Ramírez de Arellano et al., 2006).

The *gag*, *pol* and *env* genes, from the 5' end - to the 3'-end, encode the major structural and non-structural proteins (**Table 1.1**). The *gag* gene encodes a polyprotein precursor whose name, $Pr55^{Gag}$, is based on its molecular weight. $Pr55^{Gag}$ is cleaved by the viral protease (PR) to the mature gag proteins matrix (MA, p17), capsid (CA, p24), nucleocapsid (NC, p7) and p6. Two spacer peptides, p2 and p1, are also generated upon $Pr55^{Gag}$ processing.

The *pol*-encoded enzymes are initially synthesized as part of a large polyprotein precursor, $Pr160^{GagPol}$ as a result of a rare frame-shifting event during the translation of $Pr55^{Gag}$. Viral protease cleaves individual *pol*-encoded enzymes, PR, reverse transcriptase (RT) and integrase (IN), from $Pr160^{GagPol}$. The envelope glycoproteins are also synthesized as a polyprotein precursor. Unlike the Gag and Pol precursors, which are cleaved by the viral PR, the Env precursor, known as gp160, is processed by a cellular protease during Env trafficking to the cell surface. The processing of gp160 results in the generation of the surface (SU) Env glycoprotein gp120 and the transmembrane (TM) glycoprotein gp41.

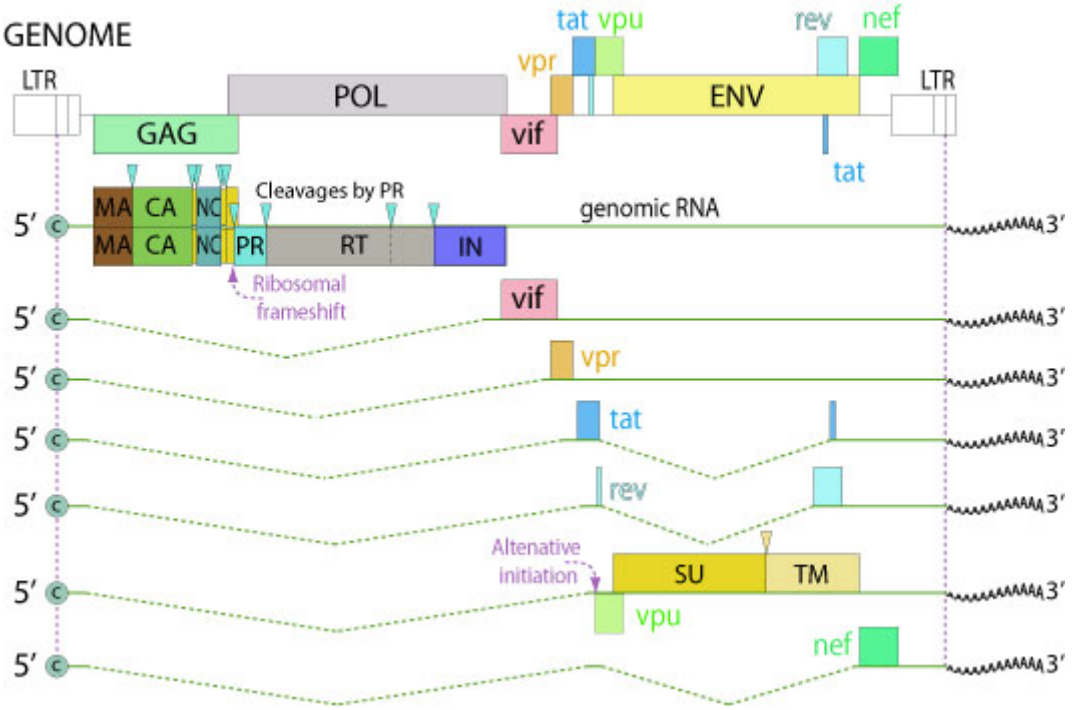



**Figure 1.5.** HIV genome structure showing the open reading frames, genomic products and their cleavage sites, as well as sites for ribosomal framshift and alternative initiation (Source: http://www.expasy.org/viralzone, Swiss Institute of Bioinformatics).

The *tat* and *rev* genes consist of two exons, with the second exon located within the *env* open reading frame. The other six HIV-1 genes have only one exon. Other accessory or auxiliary genes include *vpu*, *vif*, *vpr* and *nef*. The HIV genome also encodes several cis-acting elements, or control elements that are found adjacent to or near genes that they influence and that are involved in virion packaging, RT, RNA retention in the nucleus, RNA export and processing and transcription (Freed, 2001).

**Table 1.1.** HIV proteins, size and localisation in the virion

(Source: Kuiken, et al., 2008)

| Name | Size | Function | Localization |
|---|---|---|---|
| Gag | | | |
| MA | p17 | Membrane anchoring; *env* interaction; nuclear transport of viral core (myristoylated protein) | Virion |
| CA | p24 | Core capsid | Virion |
| NC | p7 | Nucleocapsid, binds RNA | Virion |
| | p6 | Binds Vpr | Virion |
| Pol | | | |
| PR | p15 | Gag/Pol cleavage and maturation | Virion |
| RT | p66,p51 | Reverse transcription, RNase H activity | Virion |
| RNase H | p15 | | |
| IN | p31 | DNA provirus integration | Virion |
| Env | | | |
| SU | gp120 | Binds CD4 receptors and CCR5/CXCR4 coreceptors | Plasma membrane/ Virion envelope |
| TM | gp41 | Fusogenic | |
| Tat | p16/p14 | Viral transcriptional transactivator | Primarily in nucleolus/ nucleus |
| Rev | p19 | RNA transport, stability and utilization factor (phosphoprotein) | Primarily in nucleolus/ nucleus shuttling between nucleolus and cytoplasm |
| Vif | p23 | Promotes virion maturation and infectivity | Cytoplasm (cytosol, membranes) virion |
| Vpr | p10-15 | Promotes nuclear localization of pre-integration complex, inhibits cell division, arrests infected cells at G2/M | Virion nucleus (nuclear membrane?) |
| Vpu | p16 | Promotes extracellular release of viral particles; degrades CD4 in the ER; | Integral membrane protein |
| Nef | p27-p25 | CD4 and MHC class I down-regulation (myristoylated protein) | Plasma membrane, cytoplasm, (virion) |

## 2.4 HIV life cycle

### 2.4.1 Entry

The HIV-1 life cycle is summarized in **Figure 1.6.** Viral entry into the cell is initiated by the binding of the SU glycoprotein, located on the viral membrane surface, to specific cell surface receptors. The major receptor for HIV-1 is CD4, an immunoglobulin (Ig)-like protein expressed on the surface of a subset of T cells and primary macrophages. Members of the G protein-coupled receptor superfamily of seven-transmembrane domain proteins, the α-chemokine receptor CXCR4 and the β-chemokine receptor CCR5, also function as coreceptors for HIV entry.

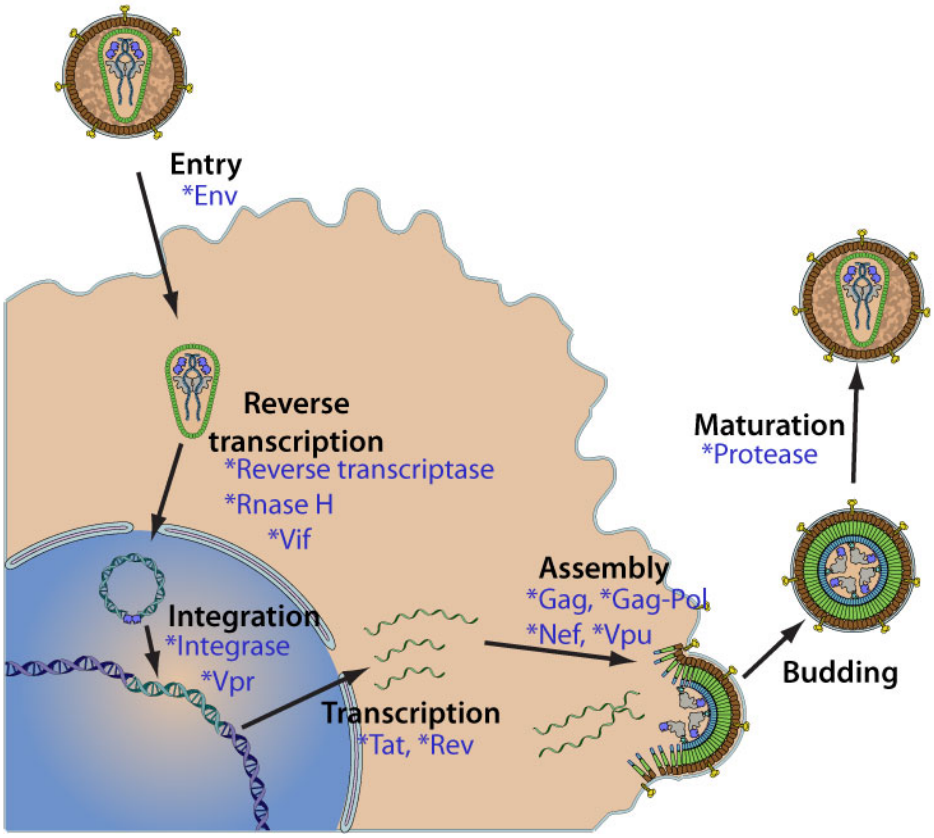



**Figure 1.6.** Replication cycle of HIV (Source: http://www.expasy.org/viralzone Swiss Institute of Bioinformatics).

The HIV envelope protein exists as a homotrimer composed of three non-covalently associated heterodimers of the gp120 (surface, SU) and gp41 (transmembrane, TM) subunits on the membranes of both HIV virions and productively infected cells (Garber et al., 2004).

The interaction between the gp120 of the virion and the CD4+ receptor and α-chemokine or β-chemokine coreceptors on the target cell induces the two heptad repeat motifs of the gp41 ectodomain to form a six helical bundle, which facilitates fusion between the viral envelope and cell membrane. The membrane fusion reaction between the lipid bilayers of the viral envelope and the host cell plasma membrane enables the viral core to gain access to the cytoplasm of the target cell and is thus central to the infection process (Freed, 2001).

### 2.4.2 Reverse transcription

HIV, as a member of the *Retroviridae* family, utilizes reverse transcriptase to convert genomic RNA into complementary DNA as part of its replication process. HIV-1 RT is an asymmetric heterodimer composed of p66 and p51 subunit. The p51 polypeptide is derived from the p66 by proteolytic cleavage of its C-terminal RNase H domain. The p66/p51 HIV-1 RT heterodimer contains one DNA polymerization active site and one RNase H active site, which both reside in the p66 subunit at spatially distinct regions, while the p51 subunit fulfils a structural function.

Cellular $tRNA_{lys3}$ binds to the genome of HIV-1 at the primer binding site (PBS) to initiate the synthesis of the minus (–) DNA strand, and simultaneously the RNA strand is digested by RNase H. The PBS is near the 5′ end of the genome and degradation of the RNA by RNase H allows DNA synthesis to be transferred to the 3′ end of the RNA. After strand transfer, (–) strand synthesis continues accompanied by RNase H degradation of the RNA genome which does not proceed to completion. The purine-rich polypurine tract (PPT) is resistant to RNase H cleavage and serves as the primer for plus (+) strand synthesis. The PPT sequence is just 5′ of U3. Removal of the PPT primer by RNase H defines the left end of the upstream long terminal repeat

(LTR), which, together with the downstream LTR, is the substrate for the viral integrase enzyme that inserts the linear viral DNA in the host genome (Sarafianos, et al., 2001; Sarafianos et al., 2004).

### 2.4.3 Integration

The ability of lentiviruses to infect nondividing cells stems from lentiviral nucleophilic proteins that allow the viral cDNA to translocate through the nuclear envelope in a nondividing cell. The IN catalyzes the insertion of the linear, double-stranded viral DNA into the host cell chromosome. The integration of the viral DNA into the host cell chromosome generates a “provirus”, which behaves essentially as a cellular gene (Craigie and Bushman, 2012; Freed, 2001).

### 2.4.4 Transcription

The transcription of the provirus into both the genome and mRNA molecules completes the HIV replication cycle. The HIV-1 LTR is the site of transcriptional initiation and contains *cis*-acting elements required for RNA synthesis. The LTR comprises three regions: U3, R and U5 and transcription initiates at the U3/R junction while U3 contains an array of elements that direct the binding of RNA polymerase II (RNA Pol II) to the DNA template including a TATA element, to which transcription factor IID binds.

Rev and the RRE, a large, highly structured RNA element located in the *env* gene and present in all unspliced and partially spliced HIV-1 RNAs, cooperate to transport unspliced and partially spliced RNAs from the nucleus to the cytoplasm. Unspliced viral RNAs function as the mRNAs for the Gag and Gag-Pol polyprotein precursors, and are packaged into progeny virions as genomic RNA, and partially spliced mRNAs encode the Env, Vif, Vpu, and Vpr proteins (Freed, 2001).

### 2.4.5 Assembly

The assembly of lentiviruses, including HIV-1, takes place at the plasma membrane of the infected cell. The final step in the process of virus assembly

and release involves the pinching off, or budding, of the virus particle from the host cell plasma membrane. During, and immediately following the release of the virus from the plasma membrane, the viral PR cleaves the Gag and GagPol polyprotein precursors to generate the mature Gag and Pol proteins. Following cleavage, CA forms a conical shell around the RNA/protein complex inside the core (Freed, 2001; Sundquist and Kräusslich, 2012).

## 3. HIV-1 diversity

HIV is characterized by extreme genetic variability. The first cause of this extraordinary genetic heterogeneity is the result of the multiple introductions of genetically diverse simian viruses into human (Hahn et al., 2000). A system that is currently widely used for classifying HIV-1 and HIV-2 into a number of subtypes is based primarily on genetic sequences coding for the envelope (*env*) and structural (*gag*) proteins and methods to infer the phylogenetic relationships between them (Hu, et al., 1996).

The following criteria were laid down for the current classification system into subtypes: subtypes are approximately equidistant from one another in *env*; the *env* phylogenetic tree is for the most part congruent with the *gag* phylogenetic tree; and two or more samples are required to define a new sequence subtype (Janssens et al., 1997). The identification of these subtypes and CRFs is important in epidemiological tracking and in the understanding of the ever-changing epidemic (Skar et al., 2011).

### 3.1 HIV-1 groups M, N, O and P

HIV-1 and SIVcpz lineages are intermingled within the evolutionary tree, defining three distinct groups of HIV-1, M, N, O and P, each of which must have arisen from a separate cross-species transmission (**Figure 1.7**). African primates are infected by simian immunodeficiency viruses, and the two major viral types infecting humans, HIV-1 and HIV-2, represent zoonotic transmissions from two different primates, *Pan troglodytes troglodytes* (chimpanzees) (Gao, et al., 1999) and *Cercocebus atys* (sooty mangabeys) (Gao, et al., 1992), respectively. HIV-1 groups M and N originate directly, but independently, from SIVcpz found in the chimpanzee *Pan troglodytes*

*troglodytes* in West-Central Africa. HIV-1 group O and group P are related to SIVgor identified in gorillas (*Gorilla gorilla gorilla*) in Cameroon (Gao et al., 1999; Hahn et al., 2000; Hemelaar, 2012; Keele, et al., 2006; Plantier, et al., 2009; Takehisa, et al., 2009; Vallari, et al., 2011; Van Heuverswyn, et al., 2006).

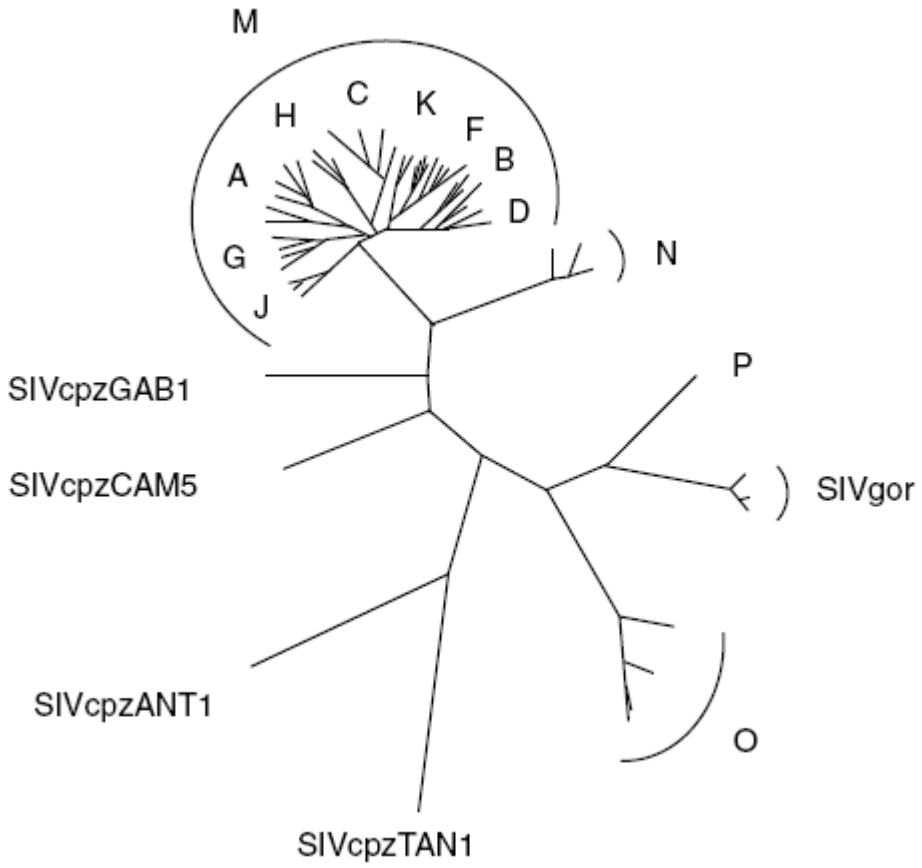



**Figure 1.7.** Phylogenetic tree of SIV and HIV-1. Phylogenetic tree based on polymerase sequences of reference sequences of SIV and HIV-1.
(Source: Hemelaar, 2012).

### 3.2 HIV-1 subtypes, CRFs and URFs

HIV-1 groups have spread very unevenly and the global AIDS pandemic is largely the result of HIV-1 group M. Since its advent in the human population, the HIV-1 group M has undergone rapid diversification. The M group is classified into various subtypes, inferred from the “starburst” radiation from the root of the group M evolutionary tree, which seem to have arisen from incomplete sampling, local founder effects, as well as the exponential growth

of the number of infected individuals (**Figure 1.8**) (Hahn et al., 2000; Rambaut et al., 2001; Sharp, 2002).

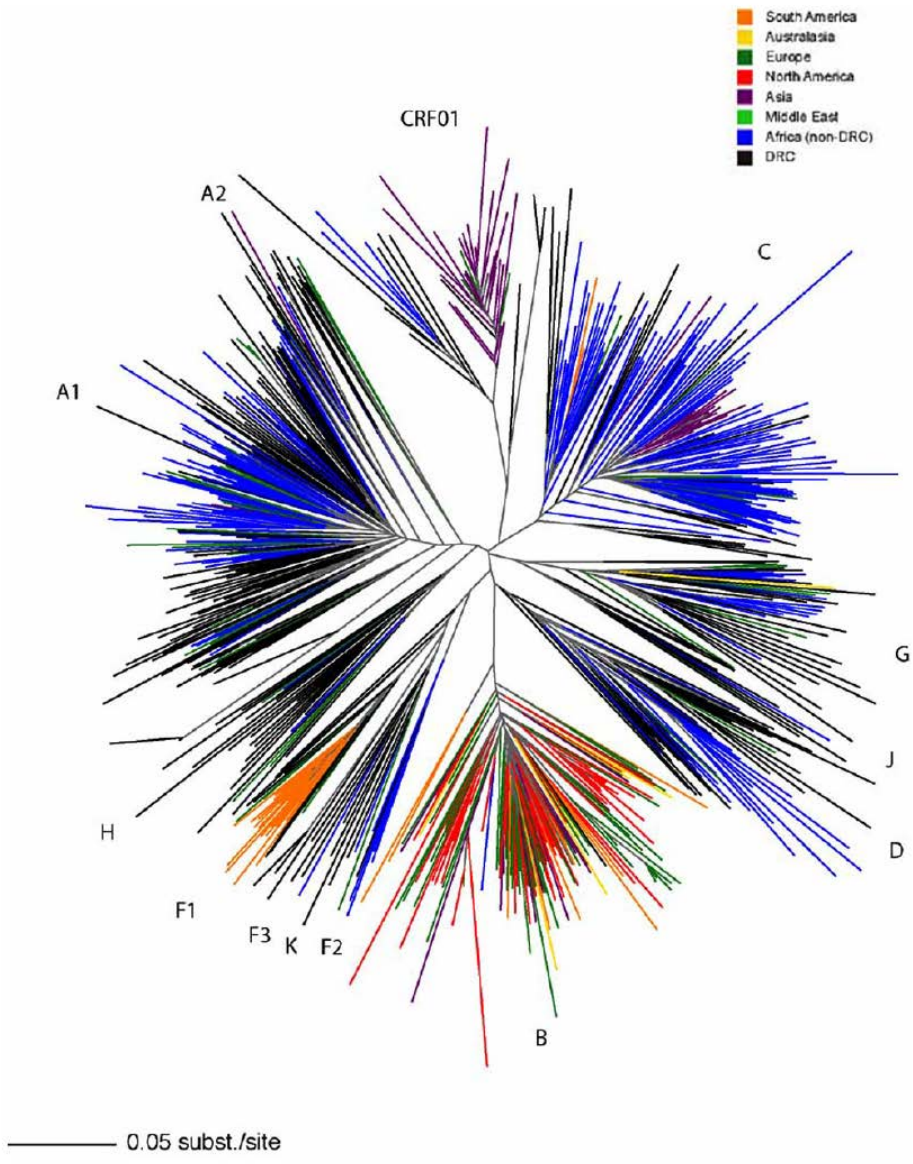



**Figure1.8.** HIV-1 group M diversity. Phylogenetic relationships among subtypes, sub-subtypes and CRFs based on *env* sequences (V2-V5) of HIV-1 isolates sampled worldwide. Strains are color-coded according to their geographical origin. DRC samples are coded separately since they are more diverse than all the strains sampled worldwide. (Source: Butler et al., 2007).

Founder effects involved in the spread of HIV-1 infection have resulted in the uneven distribution of M group strains among geographic areas and exposure risk populations. This accounts for the fact that HIV-1 strains tend to exhibit specific associations with particular geographic regions and/or modes of

transmission (Gifford, et al., 2007). HIV-1 phylogenetic classification is summarized in **Table 1.2.** Subtypes are genetically defined lineages that can be resolved through phylogenetic analysis of the HIV-1 M group as well-defined clades, or branches, in a tree (Gaschen, et al., 2002). The HIV-1 subtypes are approximately equidistant evolutionary relationship, except for subtypes B and D (Sharp et al., 1999).

Currently, strains belonging to the same subtype can differ by up to 20% in their envelope proteins, and between-subtype distances can soar to 35% (Gaschen, et al., 2002). Within a subtype, variation at the amino acid level is in the order of 8-17% but can be as high as 30%, whereas variation between subtypes is usually between 17% and 35% but can be up to 42%, depending on the subtypes and genome regions examined (Hemelaar, 2012). These subtypes differ from one another by an average nucleotide percentage distance of 27%. And within each subtype, there is a plethora of HIV-1 variants that represent expressions of intra-subtype diversity with average nucleotide distances approximately 11% (Janssens et al., 1997).

Recombination occurs frequently, and a CRF carries sections of two or more subtypes in a mosaic genome; a recombinant lineage is designated a CRF when related forms are found in multiple epidemiologically unlinked individuals (**Table 1.2**).

**Table 1.2.** Phylogenetic classifications of HIV-1 (Source: Taylor et al., 2008).

| Classification | Definition | Examples |
|---|---|---|
| Subtypes or clades | Genetically related HIV-1 strains that are essentially phylogenetically equidistant, generating a starlike, rather than a treelike, phylogeny. | Subtypes A, B, C, D, F, G, H, J, and K are currently known; A through D are highly prevalent, others have low prevalence and limited geographic distributions |
| Sub-subtypes | Distinct lineages within a subtype; genetic distance between sub-subtypes is smaller than that between subtypes | Subtypes A and F are subdivided into subsubtypes A1 through A4 and F1 and F2, respectively; mostly these circulate in Central and West Africa |
| Intersubtype recombinant forms | Mosaic strains with segments from two or more subtypes alternating across the genome | Common in mixed-subtype epidemics; thought to result from infection of a person with more than one HIV-1 subtype |
| Circulating recombinant forms | Specific recombinant forms that are spreading in a population; new forms are defined when three people without direct epidemiologic linkage are found to be infected; the assigned name reflects sequence of discovery and subtype composition, with “cpx” indicating forms containing three or more subtypes | Currently, 51 forms are described; CRF01_AE and CRF02_AG are found principally in Southeast Asia and West Africa, respectively; others have more limited distributions |
| Unique recombinant forms | Intersubtype recombinant forms recovered from only a single person | Hundreds of forms have been described on the basis of partial or complete genome sequences; their potential for epidemic spread is unknown |
| Second generation recombinants | Some CRFs have recombined further with other subtypes or CRFs giving rise to SGRs | |
| Geographically distinct lineages | Lineages, often country-specific, that are distinguishable phylogenetically; unlike subsubtypes, they are not phylogenetically equidistant within subtypes | Thai B, Indian C, West vs. East African D, and Former Soviet Union A (FSU-A) |

## 3.3 The origins and causes of HIV diversity

### 3.3.1 High rate of replication and rapid viral turnover.

The estimated average total HIV-1 production is 10.3 x $10^9$ virions per day. The minimum duration of the HIV-1 life-cycle in vivo is 1.2 days on average, and the average HIV-1 generation time – defined as the time from release of a virion until it infects another cell and causes the release of a new generation of viral particles – is 2.6 days (Hoet al., 1995; Perelson et al., 1996; Wei, et al., 1995)

### 3.3.2 The low fidelity of HIV RT

The HIV-1 reverse transcriptase enzymes are encoded by viral *pol* gene and use either a DNA or RNA template and, like all RNA polymerases, lack the 3'→5' exonucleolytic proofreading activity. Estimates of error rates during polymerization *in vitro* range from one error in every $10^2$ to one error in every $10^7$ nucleotides polymerized (Preston and Dougherty, 1996). The mutation rate of HIV-1 in this system was determined to be more or less 3.4 x $10^{-5}$ mutations per base pair per replication cycle (Mansky and Temin, 1995). Thus the misincorporation, deletion, insertion, or duplication of nucleotides arising from the reverse transcriptase's lack of 3'→5' exonucleolytic activity has been estimated to occur at a frequency of $10^{-4}$ to $10^{-5}$ (Malim and Emerman 2001). The most commonly detected mutations were base substitution mutations, G to A and C to T transition mutations, and frameshift mutations, -1 frameshifts in runs of T's and A's, (Mansky and Temin, 1995).

### 3.3.3 Viral quasispecies

The notion of viral quasispecies seeks to describe the extensive genetic variation that characterizes populations of RNA viruses. Thus viral quasispecies represent discrete ensembles of viral genomes defined in time and space (Vartanian et al., 1994). The diversity of viral quasispecies is shaped by a combination of mutation and selection forces. The main selective forces that have been proposed to drive HIV diversity are the immune response, cell tropism, and random activation of infected cells (Bonhoeffer et

al., 1995). The quasispecies model for RNA viruses predicts the existence of a replication error threshold beyond which there is a melting or total loss of sequence information. The retroviral G to A hypermutation is probably an example of this replication error. The G to A transitions within GpA may result from temporary dislocation of the primer and template strands by a single base. The two circumstances may be related by the local dCTP substrate concentration. G to A hypermutation is an example of induced mutation whereby the viral reverse transcriptase is forced into making errors by imbalances in the intracellular dCTP concentration (Vartanian et al., 1994).

### 3.3.4 Recombination

Two copies of viral RNA are encapsidated in HIV-1 virions, and the two genomes can be genetically distinct if the cell producing the virus has been infected with two or more different viruses. During the ensuing replicative cycle, the alternate use of the two templates RT produces a recombinant DNA genome that is different from that of either parental strain. Extensive studies performed in vitro have shown that recombination occurs frequently throughout the genome during HIV replication (2 to 20 events/genome/replicative cycles) and can efficiently shuffle closely linked genetic markers. *In vivo*, recombination between viral variants also occurs at a high rate, reflecting the frequent occurrence of cells dually infected by genetically distinct HIV variants (Nora et al., 2007). Genetic analyses have shown that recombination occurs mainly during minus-strand DNA synthesis, which supports these models for recombination. In order for novel recombinants to be generated, the two RNAs packaged in the virion must be genetically different (heterozygous virions). Only cells infected by more than one retrovirus (double infection) can produce heterozygous virions; therefore, how often double infection occurs affects the overall frequency of recombination (Chen, et al., 2005; Hu and Temin, 1990).

Recombination among these highly divergent strains contributed further to the extensive diversity of the virus observed in Central Africa. For example, possible recombination events were first detected when phylogenetic analyses of isolates differed by which gene or region within the same gene

was used for analysis (Hu, et al., 1996; Louwagie, et al., 1993; Nkengasong, et al., 1994; Paraskevis and Hatzakis, 1999).

## 4. Public health implications of HIV diversity

### 4.1 HIV diversity and mode of transmission

The first HIV-1 isolates were derived mostly from men who have sex with men and injecting drug users in North America and Western Europe, while HIV-1 subtype distribution among the heterosexual population in other regions of the world tended to be genetically diverse (Tatt et al., 2001). Studies of subjects from Thailand, the Netherlands, Brazil, Russia, Denmark, Argentina and Australia have demonstrated that individuals from different risk groups are infected by distinct populations of HIV variants (Learn, and Mullins, 2003). A study found that patterns of HIV-1 genetic diversity could be explained to some extent by the risk factors of MSM and IDU/IDX (Anderson et al., 2003).

A subtype might initially predominate in a specific transmission group but there is little supporting evidence to suggest that subtypes A, B, C, D and CRF01_AE are any more or less transmissible by a specific route, in a specific ethnic group or in specific cell types. In addition, such initial predominance is generally lost during an epidemic. For example, in Thailand HIV-1 subtypes B and CRF01_AE were segregated to intravenous drug users and heterosexuals, respectively, in the late 1980s but are now more uniformly mixed in the Thai population (Ariën et al., 2007). A study did not consider race, gender, or risk factors for the selection of the local controls because of a lack of evidence suggesting a correlation between these factors and specific HIV-1 DNA sequences (Metzker et al., 2002).

### 4.2 HIV diversity, rate of transmission and pathogenesis.

The uneven distribution of HIV-1 subtypes in different regions of the world does not necessarily stem from the inherent properties of the viral strains themselves but may arise from founder effects in which ne or more subtypes are introduced and consequently established in a population before other subtypes. The suggestion that virological factors such as HIV-1 subtype can

determine patterns of transmission and disease progression in different regions of the world amounts to an ‘ecological fallacy’, or inferring individual risk from apparent associations at the population level. Differences and changes in subtype distribution in different populations are ecological and do not reflect incontrovertible evidence for differences in transmissibility between the various subtypes (Hu, et al., 1996).

There were no significant differences found in the rate of disease progression or in the rate of CD4 lymphocyte decline between individuals infected with subtypes A, B, C or D. And it was found that Israeli men infected with HIV-1 subtype B and Ethiopian immigrants infected with HIV-1 subtype C displayed similar rates of CD4 lymphocyte decline. This despite the fact that the CXCR4 positive, SI phenotype, a biological phenotype that has been associated with an increased risk of rapid disease progression among individuals infected with subtype B, is rarely found among subtype C isolates (Alaeus et al., 1999; (Ariën et al., 2007). Despite the fact that the 32-base pair deletion in the CCR5 gene, which has been reported to be associated with a favourable outcome after HIV-1 infection, has an allelic frequency of approximately 10% in Caucasians and is virtually absent in Africans, it was found that there was no difference in the rate of disease progression between ethnic Africans and ethnic Swedes (Alaeus et al., 1999).

There are subtype-dependent differences in coreceptor usage in which CXCR4-positive isolates were rare in subtype C and dual tropism for CXCR4 and CCR5 was rare among subtype D isolates. While, subtype-C-infected individuals did not differ from other individuals in clinical or immunological status, there is the possibility that subtype-specific difference in the correlation between biological phenotype and CXCR4 usage may have implications for biological properties such as virulence, tissue tropism, or transmissibility (Tscherning et al., 1998). HIV-1 subtype D may be more pathogenic than subtypes A and C and recombinant viruses. The reason subtype D may be more pathogenic compared to other subtypes could be because subtype D has a dual coreceptor tropism, both CCR5 and CXCR4, unlike subtypes A and

C which only express tropism for CCR5 (Baeten, et al., 2007; Huang et al., 2007; Kiwanuka et al., 2008; Vasan, et al., 2006).

HIV-discordant couples in which the initially HIV negative partners were monogamous were found to have a statistically significant higher rate of HIV transmission for subtype A viruses relative to subtype D, which may be due to the differences in coreceptor tropism between the two subtypes (Huang et al., 2007; Kiwanuka et al 2009). There were also significant differences in the distribution of *in utero* transmission time found between HIV-1 subtype C *env,* subtype A *env,* and subtype D *env* (Renjifo et al., 2004).

### 4.3 HIV diagnostics and management

Genetic diversity poses a serious challenge for HIV-1 testing, diagnosis and monitoring (Butler et al., 2007).

#### 4.3.1 Diagnostics

The majority of HIV serological diagnostic tests were developed using HIV-1 subtype B. The third and fourth generation serological diagnostic tests can detect both antigens and antibodies. While many third and fourth generation serological tests are sensitive to HIV-2, group O and most group M subtypes, false negatives have been observed with HIV-1 subtype D. The fact that third and fourth generation tests can detect early has, however, produced varying results among clades. And it has been observed that CRF01_AE has a significantly longer “early” window period with these fourth generation serological tests than subtype B. These third and fourth generation serological diagnostic tests for the p24 antigen tests showed 98.1% sensitivity with HIV-1 subtype C (Butler et al., 2007). The fact that HIV rapid diagnostic tests (RDTs) also tend to be based on antigens from HIV-1 subtype B may impact on their ablity to detect diverse HIV-1 and HIV-2 subtypes circulating regionally, as well as within patient recombinant and dual infections. But one study found that the OraQuick HIV-1/2 RDT (OraSure Technologies, Inc., Bethlehem, PA) effective in detecting isolates from Kinshasa, DRC. A study to determine the sensitivity of 6 HIV-1/HIV-2 rapid screening tests in comparison to other HIV diagnostic tests, including a third generation EIA, using a well characterized

panel of HIV sera derived from HIV-1 groups M and O, HIV-2, as well as serial samples from two primates experimental primary SIV infection to mimic HIV-2 seroconversion samples, found that the rapid tests only became positive 2 to 8 days later than the third generation EIA (Makuwa, M. et al., 2002; Shott et al., 2012).

### 4.3.2 Management

Monitoring viral load (VL) or the quantification of HIV-1 RNA in serum or plasma is used to determine disease progression and response to treatment is important in the clinical management of HIV (Butler et al., 2007, Swanson et al., 2003). Nucleic acid-based assays for the detection and monitoring of viral load involve the amplification of a HIV-1 target region and detection using sequence-specific probes. Sequence polymorphism in primer and probe binding regions can lead to the underestimation or even failure to detect RNA from some HIV-1 subtypes, strains or isolates (Sizmann et al., 2010; Swanson et al., 2003).

There are several methods of determining VL, including real-time PCR (RT-PCR) (Amplicor, Roche), branched DNA assays (bDNA, Bayer), nucleic acid sequence based amplification (NASBA, Vironostica) and ligase chain reaction (LCR, Abbott). The fact that early VL assays were designed for HIV-1 subtype B meant, since real-time-PCR depends on specific primers, they often could not detect non-B subtypes, particularly A, CRF01_AE and F (Butler et al., 2007).

A study found that among individuals infected with CRF02-AG strains the sensitivity and average viral load values were significantly lower for the NucliSens assay (Organon Teknika, Boxtel, the Netherlands) than for the Amplicor HIV-1 Monitor (version 1.5; Roche Diagnostics, Branchburg, NJ, USA) and the Versant HIV-1 RNA (Bayer, Cergy Pontoise, France). The HIV monitor assay and the NASBA assay appear unable to accurately quantify HIV-1 RNA levels in plasma samples from many individuals infected with subtype-A, which may stem from primer mismatches (Alaeaus, et al., 1997).

The NucliSens HIV-1 QT assay (bioMerieux), NASBA, misses subtypes G and A, and may underestimate Group O (WHO/UNAIDS 2007).

### 4.4 HIV diversity and antiretroviral therapy and drug resistance

HIV-1 subtype differences do not impact the response to antiretroviral therapy.The drug susceptibility testing of subtype A, B, C, and E isolates revealed that the non-B isolates were also similarly inhibited by azidothymidine (AZT), lamivudine (3TC), didanosine (ddl), ritonavir and nevirapine (Palmer et al 1998). No association was found between HIV-1 subtype and virologic response to antiretroviral therapy in HIV-1-infected children (Jülg, and Goebel, 2005). Isolates derived from different HIV-1 subtypes, including G, C, A, J, D, and F, were all found to be susceptible to protease inhibitors (Holguín et al., 2004). HIV-1 subtype C isolates from Zimbabwe were found to be all found to be susceptible to antiretroviral drugs (Shafer et al, 1999).

HIV subtypes can take different pathways in evolving resistance to antiretroviral drugs (Grossman, et al 2004a, Johnson et al., 2010). Mutations in the HIV-1 subtype B reverse transcriptase that arise in the presence of zidovudine or stavudine, the thymidine analogue-associated mutations or TAMs, involve alternate, mutually exclusive, pathways, designated TAM1: M41L, T215Y, L210W, D67N, and TAM2: D67N, T215F, K70R, K219QEN. However, HIV-1subtype C-infected adults treated with first line ZDV/ddI-containing HAART in Botswana, the 67N 70R 215Y genotype was the dominant pattern of NRTI-associated mutations at the time of virologic failure. It is possible that the 67N, 70R, 215Y genotype represents a unique TAM pathway among HIV-1 subtype C-infected individuals treated with ZDV/ddI-containing HAART as the first line regimen (Novitsky et al., 2007). The M41L and L210W mutations, as well as the T215Y mutation, occurred at a much lower prevalence in patients infected with non-subtype B HIV-1 isolates, particularly subtypes A1, A2, C, F1, F2 and CRF06_cpx, than in HIV-1 subtype B isolates (Montes et al., 2004).

HIV-1 subtype C reverse transcriptase possesses the valine polymorphism GTG at codon 106, which occurs in 94% of subtype C isolates but only in 1.5% of subtype B variants. The GTG polymorphism favours the occurrence of a V106M, or GTG→ATG, mutation following either *in vitro* or *in vivo* exposure to efavirenz (EFV) (Brenner et al., 2003). Exposure to EFV can also select the V106M mutation in subtypes/CRFs other than C, for example CRF02_AG (Montes et al., 2004). The V106M mutation rarely appears in subtype C patients treated with NVP (Grossman, et al 2004b; Morris et al., 2003).

There was more prevalence of nucleoside analogue-associated mutations or NAMs in HIV-1 subtype B-infected patients, while NNRTI resistance-associated mutations seemed to accumulate more rapidly in HIV-1 subtype C patients, in which they display a unique pattern. However, the M184V mutation seems to occur at equal frequencies in both subtype B- and subtype C-infected patients (Grossman et al., 2004b).

The D30N mutation in the HIV-1 protease occurs much less frequently in HIV-1 subtype C than in subtype B. The D30N mutation is a primary nelfinavir mutation, in contrast to the L90M mutation, which confers cross resistance to protease inhibitors including nelfinavir, indinavir and saquinavir. The D30N and L90M mutations are rarely selected on the same isolate, and in the majority of subtype-B-infected patients developing resistance after failing nelfinavir as their first PI, the D30N mutation is selected, while the L90M mutation or other primary mutations are selected in a significantly smaller proportion (Grossman, et al 2004a, Johnson et al., 2010).

The M36I mutation is a secondary resistance mutation to protease inhibitors that occurs as a polymorphism in HIV-1 subtype C isolates. The T12S/T15V/L19I/I93L mutation pattern has been found in HIV-1 subtype C isolates from South Africa and north India, with the latter isolates also containing the additional L89M mutation (Arora et al., 2008). The mutations T69N and V75M in reverse transcriptase and L10F, K20I, L33I, and N88S in protease have been found more frequently in patients infected with

CRF01_AE than in patients with subtype B (Ariyoshi et al., 2003; Ataher et al., 2012).

### 4.5 HIV diversity and vaccine design

The rapid evolution and extreme genetic diversity of HIV represent a formidable challenge to the development of a vaccine (Garber et al., 2004). The extreme genetic diversity of HIV imposes a formidable challenge in terms of choosing a vaccine strain including whether it would be based on an isolate derived from a particular subtype selected from a geographic region for which the vaccine is intended, constructing a consensus sequence, and using an ancestral sequence reconstructed on the basis of an evolutionary model (Gaschen et al., 2002).

It has been established that serum and HIV-1 isolate neutralization serotypes do not generally correspond directly to the genetic subtypes, that is sera from patients infected with subtype A viruses does not preferentially neutralized subtype A viruses. Thus putative serum and isolate neutralization serotypes are not identical to and presently cannot be predicted by knowledge of the genetic subtypes alone. However, subtypes B and E form discrete neutralization serotypes when compared directly against each other. The envelope glycoproteins from subtypes B and E occupy opposite extremes of the antigenic diversity spectrum, which accounts for the virtual dependency of subtype to the outcome of neutralization assays that characterizes the two subtypes. Ultimately, that genetic subtypes do not directly correspond to neutralization serotypes stems from the fact that HIV-1 subtypes are defined on the basis of primary sequence, whereas virus neutralization is a much more complex process which depends on antibody interactions with epitopes that are influenced by the tertiary and quaternary structures of glycoprotein oligomers (Gaschen et al., 2002; Kostrikis et al., 1996; Moore et al., 1996; Moore et al., 2001).

There is some evidence that viral genetic subtype may also be important as the breadth and potency of humoral responses is reported to be higher in subtype C and A than in subtype B infections (Overbaugh and Morris, 2012).

The MHC-restricted host immune response, which represents a continuous selective force on pathogens, may be influenced significantly by subtype-dependent amino acid sequence variation (Kosakovsky Pond, et al 2009; Moore et al., 1996). It has been found that MHC class I epitopes containing even a single amino acid substitution could not be recognized by T cell receptors, precluding the possibility of using a single natural isolate for a vaccine protecting against either a different subtype or even many variants from the same subtype (Cao et al., 1997; Létourneau et al., 2007). Single amino acid changes may further undermine MHC class I restricted CTL responses by diminishing recognition of a given epitope by the host's CD8+ T cells through either loss of epitope binding to MHC I molecules, or preservation of binding to MHC class I but alteration of protein processing such that immunogenic epitopes are no longer presented for recognition by a T-cell receptor (Cao et al., 1997; Garber et al., 2004). However, a subset of HLA A3 positive persons infected with HIV-1 non-B subtypes revealed frequent recognition of A3-restricted epitopes that has been well described in individuals infected with clade B virus, suggesting that CTL cross-subtype reactivity among persons infected with clade B and non-B viruses may take place, resulting in MHC class I restricted CTL recognition of immunogenic epitopes despite single or even double amino acid substitutions (Cao et al., 1997).

In addition, high rates of genomic recombination allow for mixing and matching of the most favoured mutations associated with escape from host humoral and cellular immune responses, and minimisation of viral fitness costs incurred by individual escape mutations. Both mutational and recombination mechanisms of HIV diversification rely on ongoing, active virus replication and generate extensive pools of mutant viruses from which host immune responses may select for resistant variants (Garber, D. A., et al.).

### 5. Global distribution of HIV-1 subtypes

The global distribution of the different HIV-1 subtypes and recombinants and shifts in that distribution probably arose from various causes including lineage founder effects, population growth and urbanization and improved transport

links and migration (**Table 1.3.**). The actual role of the biological properties of different subtypes and recombinants in their uneven distribution has yet to be determined (Hemelaar et al., 2011; Rambaut et al., 2004).

The initial diversification of group M may have occurred within or near the DRC, where the highest diversity of group M strains has been observed and the earliest cases of HIV-1 infection (1959 and 1960) have been documented. More precisely, molecular epidemiological studies revealed that the epicenter of the HIV pandemic is situated in the area of Kinshasa, the capital city of the DRC, where the highest genetic diversity of HIV-1 M, in number of cocirculating subtypes, intrasubtype diversity, and recombinants, has been observed and about 5% of sequences reported from DRC are not clustering with any of the known HIV-1 variants. Interestingly, the geographic distribution of HIV-1M variants is also heterogeneous within the country, and between 6 and 11 HIV-1 variants cocirculate in the different regions studied. The highest genetic diversity has been observed in Kinshasa and subtype A is the most prevalent variant in almost each region, except in the extreme south where subtype C largely predominates. Subtypes D, G, and H were found mostly in the northern and western parts of the country. In Kisangani, in the northeast, subtypes A and D are relatively well represented, as in countries bordering the DRC in the northeast (Djoko et al., 2011; Keele et al., 2006; Vidal et al., 2000; Worobey, 2008).

**Table 1.3.** The global prevalence and geographic location of the 11 most prevalent HIV-1 subtypes and circulating recomninant forms (Source: Hemelaar et al., 2011).

| Subtype/CRF | Global Prevalence | Geographic Location |
|---|---|---|
| A | 12% | East and central Africa<br>Eastern Europe<br>Central Asia |
| B | 11% | Amercas<br>Western Europe<br>Oceania<br>Middle East and North Africa<br>East Asia |
| C | 48% | Southern Africa<br>Ethiopia<br>East Africa<br>India<br>Brazil |
| D | 2% | Eastern Africa<br>Central Africa<br>West Africa |
| F | 1% | Widely and evenly spread worldwide. |
| G | 5% | West and central Africa |
| H, J and K | 1% | Central, southern and West Africa |
| CRF01_AE | 5% | South and south-east Asia |
| CRF02_AG | 8% | West Africa |
| Other CRFs | 3% | Widely and evenly spread worldwide |
| URFs | 4% | Widely and evenly spread worldwide |

There has been an increase in the global proportion and absolute growth of the epidemics of subtypes A, F, G, H, CRF01_AE, CRF02_AG and other CRFs, while the epidemics caused by subtypes D, J, K, CRF03_AB and URFs have decreased in size, with the concomitant reduction in their proportion of the global total (**Figure 1.7**). However, the epidemics caused by subtypes B and C grew at a rate below the average, which reduced their proportion of the global epidemic even though subtype C was still responsible for the largest absolute increase in the number of infections (Hemelaar et al., 2011).

While some of the CRFs that are important in the global epidemic are found in central Africa and in particular in the DRC, for example CRF01_AE and CRF02_AG, other CRFs that have become important in the epidemic seem to have arisen due to recombination events involving subtypes that are co-circulating in a particular geographic region, for example CRF31_BC and CRF07_BC in Brazil and China, respectively (Hemelaar et al., 2011).

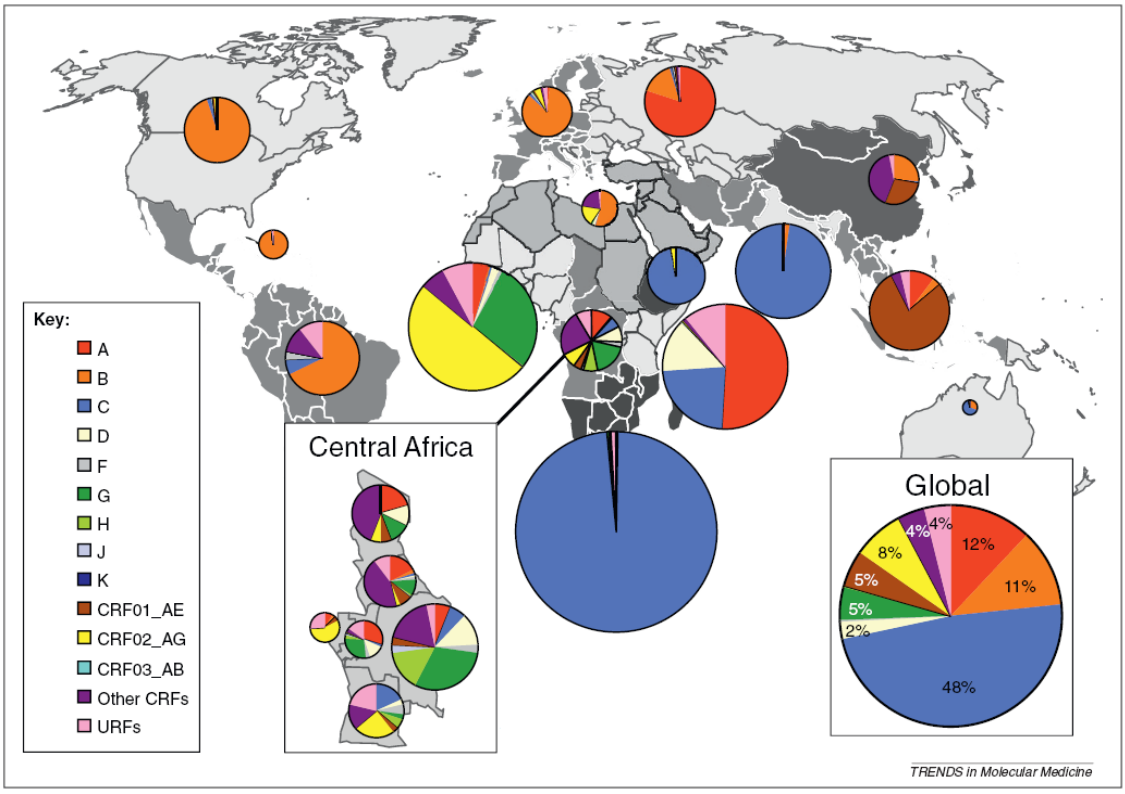



**Figure 1.9.** Global distribution of HIV-1 subtypes and recombinants. In the main figure, pie charts representing the distribution of HIV-1 subtypes and recombinants from 2004 to 2007 in each region are superimposed on the regions. The relative surface areas of the pie charts correspond to the relative numbers of people living with HIV in the regions. The colours representing the different HIV-1 subtypes and recombinants are indicated in the legend on the left-hand side of the figure. The HIV-1 subtype distributions found around the world and within Central African countries are shown in the insets of the main figure, as indicated (Source: Hemelaar et al, 2011).

## 6. Molecular epidemiology of HIV in South Africa

The explosive HIV-1 epidemic in southern Africa is characterized by very limited subtype diversity. HIV-1 subtype C is the most common subtype, accounting for the majority of HIV infections in southern Africa, while other subtypes, particularly subtypes B and D, have been identified (Abreu et al., 2008; Becker et al., 1995; Bredell et al., 2002; Engelbrecht et al., 1995; Hemelaar et al., 2011; Jacobs et al., 2009; Lihana et al., 2012; Novitsky et al., 1999; Van Harmelen et al., 1999; Williamson et al., 1995). In South Africa HIV-1 subtypes and recombinants such as F1, G, AC, AD, AG, CG, and CH have also been identified (Fish et al., 2010 Jacobs et al., 2009; Wilkinson and Engelbrect, 2009). Molecular epidemiological investigations in South Africa have, with the notable exceptions of the Free State, Limpopo and Mpumalanga, largely focused on provinces with major metropolitan centers such as Guateng, the Western Cape and KwaZulu Natal. Hardly any molecular epidemiological studies have been undertaken in the Eastern Cape, North West and Northern Cape provinces (**Table 1.4.**). HIV-1 subtype C is primarily transmitted heterosexually and affects mostly black Africans, and subtype B is associated with homosexual transmission amongst white males, which may account for the hitherto lack of instances of recombination between the two subtypes in South Africa. But sex between men could be an important factor in the HIV/AIDS epidemic in southern Africa, replicating the pattern prevailing elsewhere in the world in which men who engage in sex with other men often tend to be involved in relationships with women (Van Harmelen, et al., 1997; Van Griensven, 2007).

**Table 1.4.** The distribution of HIV-1 subtypes and CRFs by province in South Africa.

| Province | Subtype/CRF | Genomic region | Year | Reference |
|---|---|---|---|---|
| Eastern Cape | — | — | — | — |
| Free Sate | C | *pol* | 2009 | Huang, et al., 2009 |
| | B, C | *env* | 1999 | Van Harmelen, et al., 1999 |
| Gauteng | C, AC, CG | Integrase | 2010 | Fish et al., 2010 |
| | C | Near full length | 2001 | Hunt et al., 2001 |
| | A, D, G, AG | Full length | 2003 | Papathanasopoulos et al., 2003 |
| | C | *gag* and *env* | 2002 | Bredell et al., 2002 |
| | C | Full length | 2002 | Papathanasopoulos et al., 2003 |
| | C, B | *env* V3 | 1999 | Engelbrecht et al., 1999 |
| | C, B | *env* | 1999 | Van Harmelen et al., 1999 |
| | C | *env* V3-V5 | 1998 | Bredell et al 1998 |
| Kwa-Zulu Natal | C | *pol* | 2008 | Marconi et al., 2008 |
| | C | *pol* and *env* | 2003 | Gordon et al., 2003 |
| | A, D, G, AG, AC | *gag* and *env* | 2002 | Bredell et al., 2002 |
| | | *env* V3 | 1999 | Engelbrecht et al., 1999 |
| | B, C | *env* | 1999 | Van Harmelen et al., 1999 |
| Limpopo | C | *gag, pol* and *env* | 2012 | Iweriebor et al., 2012 |
| | C | *pol* | 2006 | Bessong et al., 2006 |
| | C | *env* and *gag* | 2005 | Bessong et al., 2005 |
| Mpumalanga | C, B | *env* V3 | 1999 | Engelbrecht et al., 1999 |
| Northern Cape | — | — | — | — |
| North West | — | — | — | — |
| Western Cape | A1, B, AC, AD | Near full length | 2009 | Wilkinson & Engelbrecht, 2009 |
| | B, C | *vpr* | 2009 | Romani et al., 2009 |
| | A, B, C, F1, G, CH, U | *env* V3 | 2009 | Jacobs et al., 2009 |
| | C, B | vif | 2008 | Jacobs et al., 2008a |
| | B, C, G, AG | *pol* | 2008 | Jacob et al., 2008b |
| | D | Near full length | 2007 | Jacobs et al., 2007 |
| | C, D | *env* and *gag* | 2006 | Jacobs et al., 2006 |
| | B, C | *env* | 2002 | Treurnicht et al., 2002. |
| | B, C | Full length | 2002 | Van Harmelen et al., 2001 |
| | C, B | *env* V3 | 199 | Engelbrecht et al., 1999 |
| | A, C | *env* | 1999 | Van Harmelen et al., 1999 |
| | B, C, D | *env* and *gag* | 1995 | Becker et al., 1995 |
| | A, C | *env* | 1995 | Engelbrecht et al., 1995 |
| | B, C, D | *gag* | 1995 | Williamson et al., 1995 |

### 7. Methods for detecting HIV-1 diversity.

There are several laboratory methods that are used for the detection of HIV-1 diversity. The methods encompass nucleic acid testing and serological methods and include sequencing, PCR-based methods such as heteroduplex mobility assays and serological methods (**Table 1.5.**). The sheer accumulation of HIV-1 nucleotide sequences in databases, and their preponderant use in studies on HIV-1 diversity, suggests that sequencing represents the best method for HIV-1 diversity. Serological assays are more effective in screeing populations of limited subtype diversity, for example in situations in which only one subtype predominates. In epidemics that are characterized by the co-circulation of multiple subtypes serological assays may lead to subtype misclassification because of cross reactivity. Unlike serological assays, PCR-based methods for the detection HIV-1 diversity such as heteroduplex mobility assays have been more effective in detecting HIV variants from in countries where extreme subtype diversity occurs. The use of partial sequences from different regions of the HIV genome such as *gag* and *env* has allowed prevalence of recombinants to be determined (McCutchan, 1999; Tatt et al., 2001).

**Table 1.5.** HIV-1 subype classification schemes (Source: Tatt et al., 2001).

| Method | Technique | Advantages | Disadvantages |
|---|---|---|---|
| Sequencing | Sequencing all or part of the genome | Accurate<br>Recombinant identification indicates transmission events<br>Prediction of antiviral resistance | Labour intensive and time consuming<br>Expensive |
| PCR-based | Heteroduplex mobility assay | Accurate<br>Inexpensive<br>Applicable in areas of multiple HIV-1 subtypes | Reference sequences and primers may require updating<br>May not detect diverse/novel subtypes or CRFs |
| | Restriction fragment length polymorphism | Inexpensive<br>High Specimen throughput | Applicable only in areas limited subtype diversity<br>May not detect CRFs |
| | Oligonucleotide probe hybridization | Inexpensive<br>High throughput | Subtype diversity affects probe hybridization<br>Lacks specificity in areas of multiple subtype |
| | Combinatorial DNA melting assay | Discrimination between multiple subtypes<br>Reference sequences can be tailored to a specific population | Reduced performance in areas of multiple subtype<br>Reference sequences require continual updating |
| NASBA-based | Molecular beacons and oligonucleotide probes | Real-time results | Requires probes for all subtypes |
| Serological assays | Peptide-based enzyme immunosorbent assays | Inexpensive<br>High specimen throughput<br>Relatively simple to perform | Cross-reactivity in areas of multiple subtypes<br>Applicable only in areas of limited subtype diversity<br>Does not detect CRFs |

## 8. Transmitted drug resistance

While combination antiretroviral therapy can suppress HIV replication to undetectable levels to produce with concomitant significant clinical outcomes, suboptimal suppression HIV replication can result in the emergence of drug resistant virus strains. Expert panels from the United Sates of America Department of Health and Human Services, the International Aids Society -

United States, and the EuroGuidlines Group recommend the use of drug resistance testing for resistance mutations (DRMs) associated with reduced viral response to antiretroviral treatment in the management of HIV infected patients. HIV isolates that have acquired mutations conferring reduced susceptibility to antiretroviral drugs can be can be transmitted, potentially limiting options for first line therapy in untreated individuals. Surveillance of transmitted HIV drug resistance (HIVDR) mutations therefore has profound public health implications (Gifford et al., 2009; Shafer et al., 2008).

To ensure uniformity and consistency between the disparate HIV drug resistance surveillance programmes underway in different countries and regions a standard list of mutations to characterize the epidemiology of transmitted drug resistance has been developed. A standard list of mutations affords comparison of the prevalence of transmitted resistance from different times and regions and allows meta-analyses of surveillance data collected by different groups at different times. The SDRM list should be recognized as causing or contributing to drug resistance – defined as being present on three or more of five expert lists of drug resistance mutations, the mutations should be non-polymorphic and should not occur at highly polymorphic positions, the mutation list had to be applicable to the eight most common HIV-1 subtypes, and the list should be parsimonious, excluding mutations resulting exceedingly rarely from drug pressure. While phylogenetically-linked stably transmitted resistant variants, for example the A62V NRTI mutation found in HIV-1 subtype A isolates, may also adversely impact treatment outcomes and exacerbate public health challenges arising from transmitted drug resistance, they may also bias SDRMs lists (Bennett et al., 2009; Shafer et al., 2008).

### 9. An overview of phylogenetic analysis

A phylogenetic tree is a diagrammatic representation of the evolutionary relationship between different nucleotide or amino acid sequences. A phylogenetic tree is composed of branches and nodes (**Figure 1.10.**). Branches, which are also called edges, connect nodes; a node is the point at which two, or more, branches diverge. Branches and nodes can be internal or external. External nodes are also designated terminal nodes. An internal node

corresponds to the hypothetical last common ancestor (LCA) of everything arising from it. External nodes represent entities that are being compared, and they are called operational taxonomic units (OTUs), and internal nodes represent ancestral units, which are hypothetical, the objective is to group extant units. Terminal nodes correspond to the sequences, or operational taxonomic units or OTUs, from which the tree was derived (Baldauf, 2003). The branching pattern – that is, the order of the nodes – is called the topology of the tree (Vandamme, 2003).

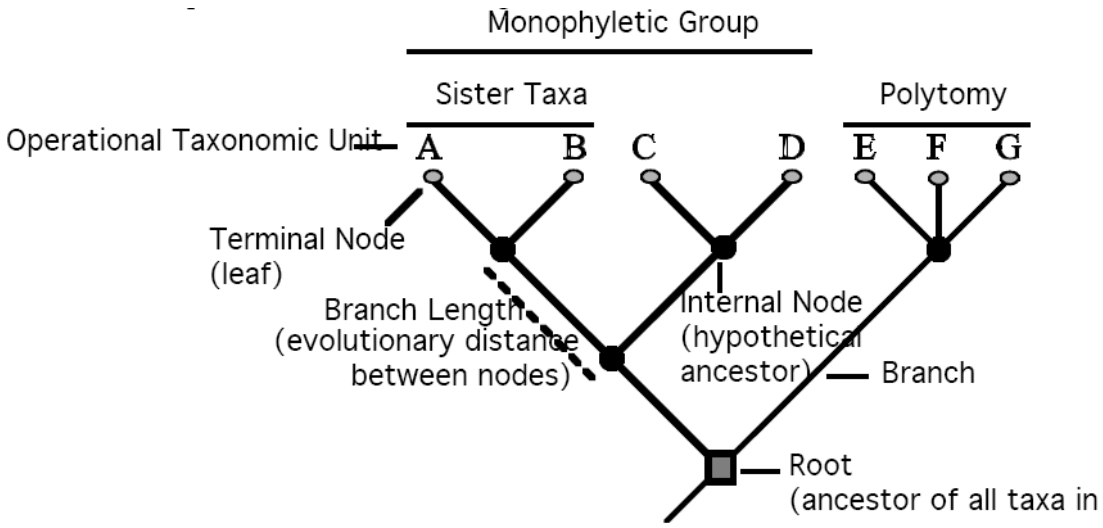



**Figure 1.10.** Example of a phylogeny with included terminology. Sister taxa are two taxa derived from a common ancestral node, they are each other's closest relatives; a monophyletic group includes an ancestor with all its descendents; a polytomy is an internal node with more than 2 immediate descendants, representing either simultaneous divergence or an unresolved node (Source: Egan and Crandall, Unpublished).

Trees can be rooted or unrooted (**Figure 1.11.**). Rooted trees have a time direction whereas unrooted trees do not. Rooted trees have a root that denotes common ancestry of all OTUs under study, and unrooted trees only specify the evolutionary relationship among taxa and not their evolutionary paths. The branching pattern of a tree is the tree's topology, and branch length reflects the amount of difference that has occurred along a branch (Baldauf, 2003). Most phylogenetic methods produce unrooted trees. This is

because they detect differences between sequences, but have no means to orient residue changes relatively to time.

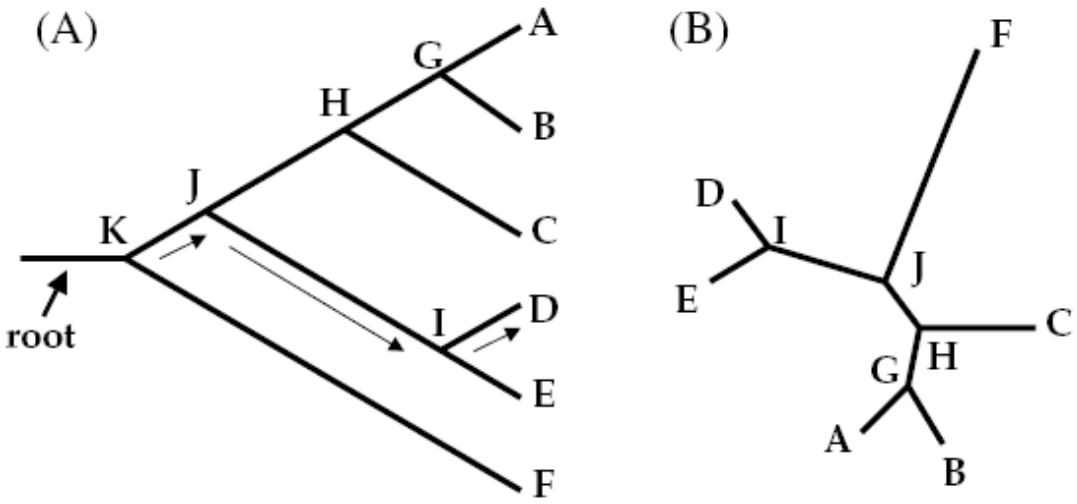



**Figure 1.11.** Examples of a rooted (A) and unrooted (B) trees (Source: Vandamme, 2003).

## 10. HIV-1 molecular phylogenetic methods

### 10.1 Assembling a DNA sequence dataset

Phylogenetic analysis to infer evolutionary relationships between organisms or genes involves the use of homologous sequence data, which can be obtained from the laboratory bench using molecular biology techniques or retrieved from sequence databases such as the LANL HIV Sequence Database. An out-group sequence is also included to root the tree, indicating which nodes in the tree are the oldest, and providing indications about ancestral sequence states or ancestral descendent relationships (San Mauro and Agorreta, 2010).

### 10.2 Multiple sequence alignment

Phylogenetic reconstruction methods require a set of aligned sequences. The alignment is the statement of homology, that is, shared ancestry from which historical inferences are made. Sequences may diverge from each other, *indels* may accumulate, and gaps need to be inserted into sequences to increase their similarity. The alignment is an arrangement of the sequences into a matrix so that the character states at each given position, column of the matrix, are related to each other by descent from a common ancestral residue, that is there is positional homology. There are two fundamental ways of treating data; as distances or as discrete characters. Distance methods first convert aligned sequences into a pairwise distance matrix, which is then used in a matrix to build a tree. Discrete methods consider each nucleotide site, or some function of each site, directly (Egan and Crandall, Unpublished; San Mauro and Agorreta, 2010).

### 10.3 Nucleotide substitution models

The most widely and generally useful nucleotide substitution models are those from the General Time-Reversible (GTR) family. Within this family we find all the standard nucleotide substitution models (JC69, K80 or K2P, F81, HKY85, TN93, GTR). The Jukes-Cantor model (JC69) is the simplest, assuming that equilibrium nucleotide frequencies are equal and that any nucleotide can change to any other with equal probability (**Figure 1.12.**) (Liò and Goldman, 1998) The Kimura 2-parameter model (K2P) allows for differences in transition and transversion rates while keeping equal base frequencies. The Felsenstein 1981 model (F81) allows some bases to be more common than others while keeping substitution probabilities equal. F81 assumes that base frequencies are similar across sequences. Combining F81 and K2P gives the Hasegawa, Kishino, and Yano model (HKY85) which allows variation in base frequencies and transition/transversion bias. The general time reversible model (GTR) builds on HKY85 by permitting each of the six substitution pairs to have different rates. GTR contains the other models nested within it (Liò and Goldman, 1998; Leitner et al., 1997)

## 10.4 Major methods for estimating phylogenetic trees

### 10.4.1 Distance methods: Neighbor-joining method

The neighbor-joining method is an algorithm that uses genetic distances to construct a phylogeny by the sequential addition of taxa (Rambaut et al., 2004). These are pairwise distance methods based on the assumption that dissimilarity between two sequences is directly related to their phylogenetic relationship. Such dissimilarity arises from the number of changes that have occurred along the branches, i.e. the evolutionary distance. In the neighbour-joining method, the DNA or amino acid sequences are first converted into a distance matrix that is then used to reconstruct a phylogenetic tree (San Mauro, and Agorreta, 2010).

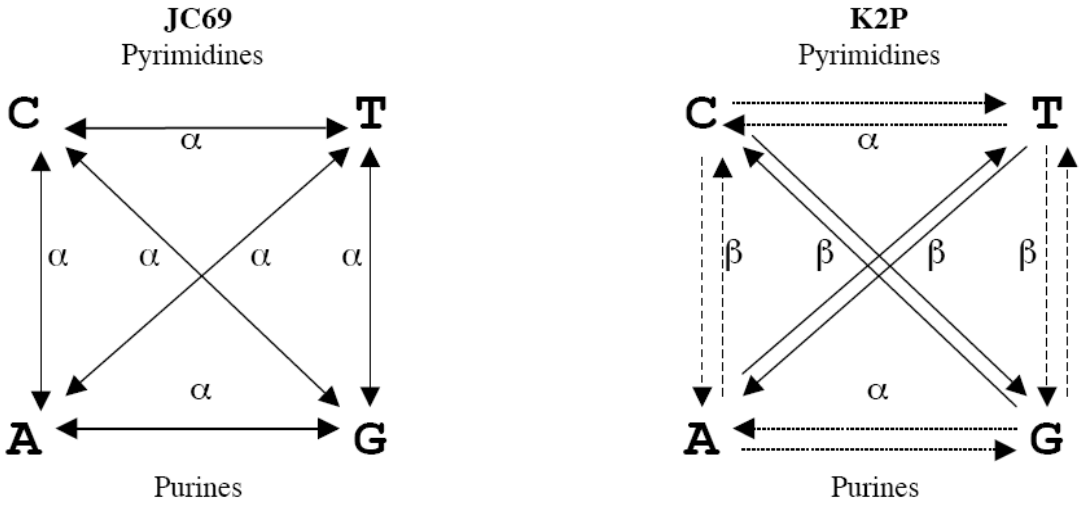



**Figure 1.12.** The Jukes-Cantor model (JC69) and Kimura 2-parameter model (K2P). α =transition rate; β = transversion rate (Source: Egan and Crandall, unpublished).

These are pairwise distance methods based on the assumption that dissimilarity between two sequences is directly related to their phylogenetic relationship. Such dissimilarity arises from the number of changes that have occurred along the branches, i.e. the evolutionary distance. Distance methods comprise both clustering methods such as Neighbour-Joining, and optimality methods such as minimum evolution. Distance methods comprise both

clustering methods such as neighbour-joining, and optimality methods such as minimum evolution. Clustering methods were originally developed to detect similarities rather than to estimate evolutionary relationships. In minimum evolution, the optimality criterion is the sum of branch lengths optimized according to the least-squares criterion above (the minimum evolution score) (San Mauro and Agorreta, 2010).

### 10.4.2 Maximum likelihood methods

Maximum likelihood methods select the tree that has the highest probability of explaining the sequence data, under a specific model of substitution, that is, changes in the nucleotide or amino-acid sequence (Rambaut, Posada, Crandall, Holmes, 2004). In the context of molecular systematics, the likelihood of a phylogenetic tree is the probability of observing the data, or the set of sequences being analyzed, given the tree and the model of evolution. The likelihood of a site is the probability of the observed states at that site given all the possible combinations of states at the internal nodes of the tree, that is, ancestral states. The likelihood of a tree is the product of the likelihoods for each site of the alignment. The great advantage of maximum likelihood is that it allows the inference of phylogenetic trees using complex models of sequence evolution, including the ability to estimate model parameters, thus allowing simultaneous inference of patterns and processes of molecular evolution, and provides a powerful statistical framework for hypotheses testing (San Mauro and Agorreta, 2010).

### 10.4.3 Bayesian inference

Bayesian inference focuses on the quantity known as the posterior probability, defined as the probability of some hypothesis conditional on the observed data. The posterior probability is proportional to the product of the likelihood of the data, given that the hypothesis is correct and the prior probability of the hypothesis before any data have been collected. In Bayesian phylogenetics, parameters such as the tree topology, branch lengths, and substitution parameters, are modeled as probability distributions. Using Bayes's theorem, the posterior probability of any of one of these parameters may be expressed as the marginal distribution of those remaining. Solving analytically for the

posterior probability requires the integration of the likelihood function over all possible values of the remaining parameters, which is effectively intractable for even moderately complex problems. Modern Bayesian methods use Markov chain Monte Carlo methods to approximate this integration by simulating draws from the joint posterior distribution of all model parameters. Posterior probabilities for the parameters of interest are calculated using the Markov chain samples. For example, the posterior probability of a tree or bipartition in a tree is determined simply by examining the proportion of all of the Markov-chain samples that contain the topological bipartition of interest (Alfaro et al., 2003).

## 10.5 Confidence Assessment

### 10.5.1 The Bootstrap

The nonparametric bootstrap procedure is commonly used for estimating nodal support under traditional methods of phylogenetic inference. The bootstrap procedure re-samples the original data set to create a new data set by choosing columns of data from the original data matrix at random with replacement until a new data matrix is created that has the same sequence length as the original. Then a tree is estimated from this re-sampled data set. This procedure is repeated multiple times (typically 100 for ML and 1000 or more for MP, ME, and NJ) to achieve reasonable precision (Felsenstein, 1985).

### 10.5.2 Posterior Probability

Posterior probability has a straightforward interpretation as the probability that a particular monophyletic group is correct.

## 11. Aims and objectives

HIV diversity may have implications for diagnosis, pathogenesis, transmission, clinical management and vaccine development. Phylogenetic analysis of HIV sequence diversity has allowed vital insights into the origin, evolution and spread of HIV, which suggests it is imperative to maintain HIV molecular epidemiology surveillance. Molecular epidemiological investigations in South Africa have, with the notable exceptions of the Free State and Limpopo, largely focused on provinces with major metropolitan centers such as Guateng, the Western Cape and KwaZulu Natal. Hardly any molecular epidemiological studies have been undertaken in the Eastern Cape, Mpumalanga, North West and Northern Cape provinces.

### 11. 1 Aim

The aim of the project was to investigate HIV-1 subtype diversity using different samples obtained from a cohort in Bushbuckridge, Mpumalanga between February and July 2009**.**

### 11.2 Objectives

To achieve the aim of the study, the following objectives were set:

1. Viral RNA isolation
2. Amplification of partial *gag*, *pol* (PR/RT and IN) and env *gp41* regions
3. Sequencing of amplified fragments
4. Molecular characterization of sequenced fragments using online tools
5. Phylogenetic analysis using MEGA software.

# CHAPTER TWO

# MATERIALS AND METHODS

# CHAPTER TWO

# MATERIALS AND METHODS

## 2.1 Patient cohort

Samples were obtained from Bushbuckridge in Mpumalanga and kindly provided for this study by Dr E Vardas. The Prevalence, Incidence and Molecular Epidemiology (PIME) study has been granted ethical approval by the Provincial Research and Ethics Committee, Department of Health and Social Services, Mpumalanga Provincial Government. The Human Research Ethics Committee (HREC) at the Faculty of Medicine and Health Sciences of Stellenbosch University also approved the study.

## 2.2 Reagents and equipment used

The reagents and equipment used are listed in the **Table 2.1** and **Table 2.2** respectively. Also provided are lists of the software used and databases (**Tables 2.3 and Table 2.4**). A diagram of the the methods used are summarized in **Figure 2.1.**

**Table 2.1.** Reagents used in the experiments.

| Product | Company |
|---|---|
| QIAamp MinElute Virus Spin Kit | Qiagen, Dusseldorf, Germany |
| Access RT-PCR | Promega Madison, WI, USA |
| Go*Taq*® Flexi DNA Polymerase | Promega Madison, WI, USA |
| dNTPs | Promega Madison, WI, USA |
| Nuclease-free Water | Promega Madison, WI, USA |
| Agarose | Whitehead Scientific |
| Ethidium Bromide | Promega Madison, WI, USA |
| 6x Blue Orange Loading Dye | Promega Madison, WI, USA |
| Wizard SV Gel and PCR Clean-up Kit | Promega Madison, WI, USA |
| 1 kb DNA Ladder | Promega Madison, WI, USA |
| ExoSAP-IT® | USB Cleveland, Ohio |
| QIAquick PCR Purification Kit | QIAGEN, Dusseldorf, Germany |
| Big Dye Terminator v 3.1 Cycle Sequencing Kit | Applied BioSystems, CA, USA |
| %x Sequencing Buffer | Applied BioSystems, CA, USA |
| Big Dye X Terminator Purification Kit | Applied BioSystems, CA, USA |

**Table 2.2.** Equipment used in the experiments.

| Equipment | Supplier Location |
|---|---|
| Eppendorf Centrifuge 5417C | Eppendorf, Hamburg, Germany |
| Eppendorf Centrifuge 5415D | Eppendorf, Hamburg, Germany |
| GeneAmp® 9700 PCR Thermocycler | Applied BioSystems, California, USA |
| NanoDrop ND 1000 Spectrophotometer | NanoDrop Technologies Inc., Delaware, USA |
| Vortex Mixer VM 300 | Gemmy Industrial Corp, Taipei, Taiwan |
| QIAcube | QIAGEN, Hilden, Germany |
| UVItec Gel Documentation System | UVItec Limited, Cambridge<br>United Kingdom |
| Hoefer EPS 2 A 200, Power Pack | Pharmacia Biotechnologies, CA, USA |
| ABI 3130xl Automated Sequencer | Applied BioSystems, CA, USA |

**Table 2.3.** Software packages used in the analysis of sequences.

| Software package | Reference / or Licensed Company |
|---|---|
| Sequencher 4.8 | Gene Codes Corporation, Ann Arbor, MI, USA |
| Clustal X | Thompson et al., 1997 |
| BioEdit v5.0.9 | Hall, 2003. Ibis Biosciences |
| MEGA v5.0 | Tamura et al., 2011 |
| jpHMM | Spang et al., 2002 |
| REGA v 3.0 HIV subtyping tool | De Oliveira et al., 2005 |

**Table 2.4.** Databases used in the analysis of sequences.

| Databases | Web Address |
|---|---|
| Los Alamos National Laboratory HIV Sequence Database | http://www.hiv.lanl.gov |
| Stanford HIV RT and Protease Sequence Database | http://hivdb.stanford.edu |
| BioAfrica | http://www.bioafrica.net |

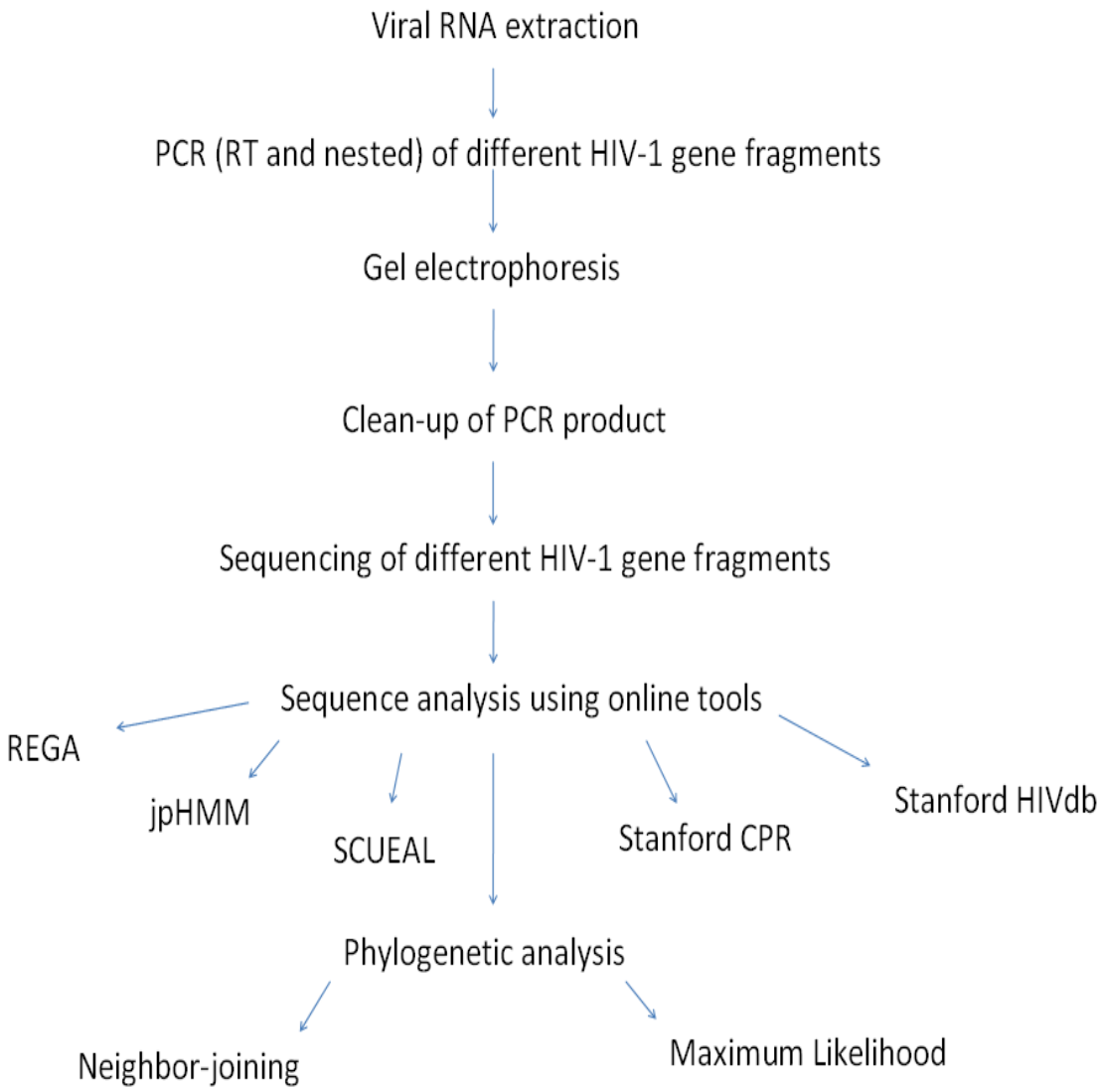



**Figure 2.1.** A flow diagram of the methods used in the study.

### 2.3 Viral RNA extraction.

HIV RNA was extracted from patient plasma samples using a QIAamp MinElute Virus Spin Kit in a QIAcube automated extractor (QIAGEN, Dusseldorf, Germany) at the Stellenbosch University Division of Medical Virology, according to the manufacturer's instructions.

### 2.4 PCR of different HIV-1 gene fragment

The reverse transcription PCR of the partial HIV-1 *gag* p24, *pol* PR p10, RT p66/51, IN p32 and *env* gp41 gene fragments was performed using the Access RT-PCR System (Promega, Madison, Wisconsin, United States of America). The Access RT-PCR System is a one-tube, two-enzyme system that uses AMV Reverse Transcriptase (AMV RT) from Avian Myeloblastosis Virus for first strand cDNA synthesis and the thermostable *Tfl* DNA polymerase from *Thermus flavus* for second strand DNA synthesis and DNA amplification. The RT-PCR reaction was carried out in an AMV/*Tfl* 5X Reaction Buffer, and the nested polymerase chain reaction (PCR) of the partial HIV-1 *gag* p24 *pol* p10, p66/51, p32 and *env* gp41 gene fragments was undertaken utilizing the Promega Go*Taq* Flexi DNA polymerase derived from *Thermus aquaticus*. The Go*Taq* Flexi DNA polymerase, which exhibits 5′→3′ exonuclease activity, is a thermostable enzyme that catalyzes the polymerazition of nucleotides into duplex DNA in a 5′→3′ direction in the presence of magnesium. The nested PCR reaction was carried in a 5X Go*Taq* Flexi Buffer. The oligonucleotide primers used in the amplification of the partial HIV-1 *gag* p24, *pol* p10, p66/51, p32 and *env* gp41 gene fragments, their locations with respect to the HXB2 reference strain, and their annealing temperatures, are in **Table 2.5**.

#### 2.4.1 Partial *gag* p24 gene fragment

Five microliters of the RNA extract was used to generate a cDNA template, which was amplified in a 50 µl reaction mixture comprising 2mM $MgSO_4$, 40pmol each of the first round PCR forward and reverse primers (Swanson, Devare, Hackett, 2003, in **Table 2.5**), 0.2 mM dNTPs and the final concentration of 0.1U each of the AMV Reverse Transcriptase and the *Tfl* DNA Polymerase. The reactions were carried out in a GeneAmp 9700 PCR System (Applied Biosystems) thermocycler according to the following cycling parameters indicated in **Table 2.6.**

One microliter of the first round DNA was amplified in a nested 50 µl reaction mixture comprising 1.5mM $MgCl_2$, 40pmol each of the nested PCR forward and reverse primers, as well as 0.2mM dNTPs and the final concentration of 1U of the Go*Taq* DNA Polymerase according to cycling parameters described in **Table 2.6**.

**Table 2.5.** Primers used in the amplification of the partial HIV-1 gene products.

| | **Primer** | **Oligonucleotide sequence** | **F/R** | **HXB2 Location** | **$T_M$ (°C)** |
|---|---|---|---|---|---|
| *gag* p24: Capsid Sequences | | | | | |
| cDNA and first round PCR | p24-7 | CCCTGRCATGCTGTCATCA | R | 1826←1844 | 55.2 |
| | p24-1 | AGYCAAAATTAYCCYATAGT | F | 1174→1193 | 56.3 |
| Nested PCR | p24-2 | AGRACYTTRAAYGCATGGGT | F | 1237→1256 | 50.0 |
| | p24-6 | TGTGWAGCTTGYTCRGCTC | R | 1703←1721 | 50.0 |
| *pol* p10 and p66/51: Protease and Reverse Transcriptase Sequences | | | | | |
| cDNA and first round PCR | PR-5′ prot -1 | TAATTTTTTAGGGAAGATCTGGCCTTCC | F | 2082→2109 | 57.0 |
| | RT-MJ4 | CTGTTAGTGCTTTGGTTCCTCT | R | 3399←3420 | 55.0 |
| Nested PCR | PR-5′ prot-2 | TCAGAGCAGACCAGAGCCAACAGCCCCA | F | 2136→2163 | 68.0 |
| | RT-NE135 | CCTACTAACTTCTGTATGTCATTGACAGTCCAGCT | R | 3300←3334 | 62.0 |
| *pol* p32 Integrase Sequences | | | | | |
| cDNA and first round PCR | poli8 | TAGTGGGATGTGTACTTCTGAAC | R | 5195←5217 | 55.0 |
| | poli5 | CACACAAAGGRATTGGAGGAAATG | F | 4162→4185 | 57.0 |
| Nested PCR | poli7 | AACAAGTAGATAAATTAGTCAGT | F | 4186→4208 | 49.0 |
| | poli6 | ATACATATGRTGTTTTACTAARCT | R | 5107←5130 | 48.0 |
| *env* gp41 Transmembrane Glycoprotein Sequences | | | | | |
| cDNA and first round PCR | JH38 | GGTGARTATCCCTKCCTAAC | R | 8346←8365 | 52.9 |
| | JH41 | CAGCAGGWAGCACKATGGG | F | 7816→7835 | 57.4 |
| Nested PCR | env27_Fc | CTGGYATAGTGCARCARCA | F | 7861→7879 | 54.9 |
| | Menv19_Rgg | AARCCTCCTACTATCATTATRA | R | 8278←8299 | 49.4 |

**Table 2.6.** Cycling conditions for the PreNested and Nested *gag* p24 PCR assays.

| **Cycling conditions of the PreNested *gag* p24 PCR assay** | | | |
|---|---|---|---|
| **Step** | **Temperature** | **Duration** | **Cycles** |
| Reverse transcription | 48°C | 45 min | x1 |
| Initial Denature Step | 94°C | 2 min | x1 |
| Denature | 94°C | 20 sec | x40 |
| Anneal | 45°C | 30 sec | |
| Extend | 68°C | 90 sec | |
| Final Extension | 68°C | 10 min | x1 |
| **Cycling conditions for the Nested *gag* p24 PCR assay** | | | |
| **Step** | **Temperature** | **Duration** | **Cycles** |
| Initial Fenature Step | 94°C | 2 min | x1 |
| Denature | 94°C | 20 sec | x40 |
| Anneal | 50°C | 30 sec | |
| Extend | 68°C | 60 sec | |
| Final Extension | 68°C | 10 min | x1 |

sec – seconds, min – minutes, °C – degrees Celsius, *gag* – Group-specific antigen gene.

### 2.4.2 Partial *pol* p10 and p66/51 gene fragment.

The primers used were described by Plantier and collaborators, 2005. The partial *pol* reverse transcription reaction mixture included 1mM $MgSO_4$, and 0.2mM of dNTPS and the final concentration of 1.1U each of the AMV Reverse Transcriptase and the *Tfl* DNA Polymerase and 40pmol each of the primers.

One microliter of the first round DNA was amplified in a nested 50 µl reaction mixture and included 0.2mM dNTPs, a final concentration of 1.25 U of the Go*Taq* DNA Polymerase and 40pmol each of the primers. The cycling conditions for the *pol* PCR are listed in **Table 2.7.**

**Table 2.7.** Cycling conditions for the PreNested and Nested reverse transcriptase-*pol* PCR assays.

| **Cycling conditions of the PreNested *pol* PCR assay** | | | |
|---|---|---|---|
| **Step** | **Temperature** | **Duration** | **Cycles** |
| Reverse transcription | 48°C | 45 min | x1 |
| Initial Denature Step | 94°C | 2 min | x1 |
| Denature | 94°C | 20 sec | x40 |
| Anneal | 55°C | 30 sec | |
| Extend | 68°C | 90 sec | |
| Final Extension | 68°C | 10 min | x1 |
| **Cycling conditions for the Nested gag p24 PCR assay** | | | |
| **Step** | **Temperature** | **Duration** | **Cycles** |
| Initial denature Step | 94°C | 2 min | x1 |
| Denature | 94°C | 20 sec | x40 |
| Anneal | 55°C | 30sec | |
| Extend | 68°C | 60 sec | |
| Final Extension | 68°C | 10 min | x1 |

sec – seconds, min – minutes, °C – degrees Celsius, *pol* -Polymerase

### 2.4.3 Partial *pol* IN p32 gene fragment.

Five microliters of the RNA extract was used to generate a cDNA template, which was amplified, in a 50 µl reaction mixture comprising 2mM $MgSO_4$, 40pmol of each of the cDNA and first round PCR forward and reverse primers (Swanson, Devare, Hackett, 2003), and 0.2mM dNTPs and the final concentration of 0.1U each of the AMV Reverse Transcriptase and the *Tfl* DNA Polymerase. The reactions were carried out in a GeneAmp 9700 PCR System (Applied Biosystems) thermocycler according to cycling parameters listed in **Table 2.8**.

Three microliters of the template DNA was amplified in a 50 µl reaction mixture comprising 1.5mM $MgCl_2$, 40pmol each of the nested PCR forward and reverse primers, as well as 0.2mM dNTPs and the final concentration of 0.05U of the Go*Taq* DNA Polymerase. The reactions were carried out in a GeneAmp 9700 PCR System (Applied Biosystems) thermocycler according to cycling parameters listed in **Table 2.8**.

**Table 2.8.** Cycling conditions for the PreNested and Nested reverse transcriptase IN PCR assays.

| **Cycling conditions of the PreNested *pol* p32 PCR assay** | | | |
|---|---|---|---|
| **Step** | **Temperature** | **Duration** | **Cycles** |
| Reverse transcription | 48°C | 45 min | x1 |
| Initial Denature Step | 94°C | 2 min | x1 |
| Denature | 94°C | 20 sec | x40 |
| Anneal | 50°C | 30 sec | |
| Extend | 68°C | 90 sec | |
| Final Extension | 68°C | 10 min | x1 |
| **Cycling conditions for the Nested *pol* p32 PCR assay** | | | |
| **Step** | **Temperature** | **Duration** | **Cycles** |
| Initial Fenature Step | 94°C | 2 min | x1 |
| Denature | 94°C | 20 sec | x40 |
| Anneal | 53°C | 30 sec | |
| Extend | 68°C | 60 sec | |
| Final Extension | 68°C | 10 min | x1 |

sec – seconds, min – minutes, °C – degrees Celsius, *pol* – Polymerase IN – integrase.

### 2.4.4 PCR amplification of the partial *env* gp41 gene fragment

Five microliters of the RNA extract was used to generate a cDNA template, which was amplified, in a 50 µl reaction mixture comprising 1.5mM $MgSO_4$, 40pmol of each of the cDNA and first round PCR forward and reverse primers (Ellenberger, et al., 1999; Swanson, Devare, Hackett, 2003), and final concentration of 0.2mM dNTPs and the final concentration of 0.1U each of the AMV Reverse Transcriptase and the *Tfl* DNA Polymerase. The reactions were carried out in a GeneAmp 9700 PCR System (Applied Biosystems) thermocycler according to cycling parameters listed in **Table 2.9.**

One microliter of the template DNA was amplified in a 50 µl reaction mixture comprising 1.5mM $MgCl_2$, 40pmol each of the nested PCR forward and reverse primers, as well as final concentration of 0.2mM dNTPs and the final concentration of 1U of the Go*Taq* DNA Polymerase. The reactions were carried out in a GeneAmp 9700 PCR System (Applied Biosystems) thermocycler according to cycling parameters listed in **Table 2.9**.

**Table 2.9**. Cycling conditions for the PreNested and Nested *env* gp41 PCR assays.

| **Cycling conditions of the PreNested gp41 PCR assay** | | | |
|---|---|---|---|
| **Step** | **Temperature** | **Duration** | **Cycles** |
| Reverse transcription | 48°C | 45 min | x1 |
| Initial Denature Step | 94°C | 2 min | x1 |
| Denature | 94°C | 20 sec | x40 |
| Anneal | 58°C | 30 sec | |
| Extend | 68°C | 90 sec | |
| Final Extension | 68°C | 7 min | x1 |
| **Cycling conditions for the Nested gp21 PCR assay** | | | |
| **Step** | **Temperature** | **Duration** | **Cycles** |
| Initial Fenature Step | 94°C | 2 min | x1 |
| Denature | 94°C | 20 sec | x40 |
| Anneal | 44°C | 30 sec | |
| Extend | 68°C | 60 sec | |
| Final Extension | 68°C | 10 min | x1 |

sec – seconds, min – minutes, °C – degrees Celsius, env - envelope

### 2.5 Gel electrophoresis of PCR products

The analysis of the nested amplification of the partial HIV-1 *gag* p24, *pol* PR p10, RT p66/51, IN p32 and *env* gp41 gene fragments was performed using agarose gel electrophoresis. The voltage applied on the agarose gel generates an electric field, and DNA molecules in the wells in the gel exposed to this electric field migrate toward the anode due to the negatively charged phosphates along the DNA backbone. The size of the DNA determines the rate at which it passes through the gel, which allows an effective separation of DNA fragment-length mixtures by electrophoresis. Molecules of linear, duplex DNA (form III) travel through gel matrices at a rate that is inversely proportional to the $\log_{10}$ of their molecular weight. The molecular weight of a fragment of interest can therefore be determined by comparing its mobility to the mobility of DNA standards of known molecular weight. Eight microliters of nested PCR product was run in 0.8% agarose gels at 50V for more or less 45 minutes. Five microliters ethidium bromide was added to the agarose gel during preparation to facilitate visualization of DNA fragments, and the Promega 1kb DNA Ladder Marker was used as a molecular weight marker.

After the run the DNA bands were visualised on the UV transilluminator and photographed with the UVITEC gel documentation system.

### 2.6 Clean-up of PCR products.

The ExoSAP-IT is a method for the clean-up of PCR products that is based on Exonuclease I and Shrimp Alkaline Phosphatase and that eliminates unincorporated primers and dNTPs to ensure they do not interfere with downstream applications such as sequencing. The ExoSAP-IT cleanup reagent is added directly to the PCR product to degrade primers and dephosphorylate dNTPs that were not consumed in the reaction. The ExoSAP-IT reagent is active in commonly used PCR buffers, which precludes a change in buffers. The clean up is carried out for 15 minutes at 37°C, which is followed by a 15-minute incubation period at 80°C to completely inactivate both enzymes.

### 2.7 Sequencing of different gene fragments.

The cycle sequencing reaction of the partial HIV-1 *gag* p24, *pol* PR p10, RT p66/51, IN p32 and *env* gp41 gene fragments was carried out with the Big Dye® Terminator v 3.1 Cycle Sequencing Kit (Applied BioSystems). Twenty five nanograms of purified PCR product was sequenced in a 10 µl reaction mixture comprising 5 pmol of sequencing primer, 1.3 µl of Big Dye terminator enzyme mix (Applied Biosystems), and 2.7 µl of Half Dye (Bioline). Data on the primers used in the cycle sequencing reaction appear in **Table 2.10** (Ellenberger, et al., 1999; Lindström and Albert 2003; Plantier, et al., 2005; Swanson, Devare, Hackett, 2003). The reactions were carried out in a GeneAmp 9700 PCR System (Applied Biosystems) thermocycler according to the following cycling parameters: 25 cycles comprising 96°C for 10 seconds for denaturation, 50°C for 5 seconds for primer annealing, and 60°C for 4 minutes for extension. Extension product purification, as well as the capillary electrophoresis and autoanalysis of the extension product using the ABI 3130xl Genetic Analyzer, the Data Collection software and Sequencing Analysis Software, were carried out at the Central Analytical Facility of the Stellenbosch University. Upon completion of the capillary electrophoresis and autoanalysis, the Central Analytical Facility sent the sample files to the

Division of Medical Virology where the Sequencher 4.7 (Gene Codes Corporation, Ann Arbor., Michigan, USA) was used to assemble the trace data into contiguous fragments, which were then verified, edited and saved as a text file (.txt) for subsequent analysis.

**Table 2.10.** Primers used in the sequencing of HIV-1 gene fragments.

| Primer | Oligonucleotide sequence | Base # | F/R | HXB2 Location | $T_M$ (°C) |
|---|---|---|---|---|---|
| *gag* p24: Capsid Sequences | | | | | |
| p24-2 | AGRACYTTRAAYGCATGGGT | 20 | F | 1237→1256 | 50 |
| p24-6 | TGTGWAGCTTGYTCRGCTC | 19 | R | 1703←1721 | 53 |
| *pol* p10 and p66/51: Protease and Reverse Transcriptase Sequences | | | | | |
| HIV-AK10 | TYCCCACTAAYTTCTGTATRT | 21 | R | 3316←3336 | 51 |
| HIV-AK11 | GTACCAGTAAAATTAAARCCAG | 22 | F | 2571→2592 | 48 |
| JA217 | CTTTTATTTTTTCTTCTGTCAATG | 24 | R | 2623←2646 | 48 |
| pol1D | TCCCTCAAATCACTCTTTGGC | 21 | F | 2251→2271 | 55 |
| HIV-pol 3 rev | CTGAAAAATATGCATCCCCC | 20 | R | 2882←2901 | 51 |
| HIV-PR outer3′ prot1 | GCAAATACTGGAGTATTGTATGGAT TTTCAGG | 32 | R | 2703←2734 | 59 |
| *pol* p32 Integrase Sequences | | | | | |
| poli5 | CACACAAAGGRATTGGAGGAAATG | 24 | F | 4162→4185 | 55 |
| poli6 | ATACATATGRTGTTTTACTAARCT | 24 | R | 5107←5130 | 48 |
| poli7 | AACAAGTAGATAAATTAGTCAGT | 23 | F | 4186→4208 | 48 |
| poli8 | TAGTGGGATGTGTACTTCTGAAC | 23 | R | 5195←5217 | 54 |
| *env* gp41 Transmembrane Glycoprotein Sequences | | | | | |
| env27_Fc | CTGGYATAGTGCARCARCA | 19 | F | 7861→7879 | 56 |
| Menv19_Rgg | AARCCTCCTACTATCATTATRA | 22 | R | 8278←8299 | 48 |

### 2.7.1 Quality control of final sequences using the Los Alamos National Laboratory (LANL) HIV Sequence Databases Quality Control tool.

The LANL QC tool performs a set of tests to help identify problems with HIV nucleotide sequences. The tests include determining subtype using the Recombination Identification Programme (RIP); identifying the most similar database sequences from HIV BLAST, constructing a phylogenetic tree for each sequence or all sequences using Neighbor Treemaker; determining the number of stop codons and frameshifts using GeneCutter; and detecting hypermutations using HyperMut. RIP identifies recombination in query

nucleotide sequences by calculating dissimilarity to a background alignment in a sliding window (http://www.hiv.lanl.gov/content/sequence/QC/index.html Accessed 22 November 2012).

### 2.8 HIV analysis using online tools

#### 2.8.1 REGA 3.0

The REGA HIV-1 subtyping Tool Version 3.0 (http://bioafrica.mrc.ac.za:8080/rega-genotype-3.0.2/hiv/typingtool#/ was used to subtype the sequences. The REGA subtyping algorithm matches the query sequence with a reference alignment comprising group M subtypes A–D, F–H, J and K, before generating a phylogenetic tree using the HKY model of evolution with a gamma-distributed variation among sites. Sequences that cluster with particular subtypes represent non-recombinants, whereas sequences that do not cluster with any subtype are characterized as either CRFs; recombinant viruses or unclassified viral subtypes. Recombination is detected using bootscanning methods, and phylogenetic signal from the alignment is used to determine subtype. The output of the REGA analysis is a report comprising data for the different phylogenetic trees, the bootstrap support for each of the trees, as well as a graphic image of the boot-scanning analyses and values for the phylogenetic signal (De Oliveira et al., 2005).

#### 2.8.2 jpHMM

To detect recombinants, we used the jumping profile Hidden Markov Model (jpHMM-HIV) tool (http://jphmm.gobics.de). This tool uses a pre-calculated multiple sequence alignment of the major HIV-1 subtypes to model each subtype in the alignment as a profile HMM. Recombination prediction for a query sequence involves the most probable path through the model that generates the sequence following state transitions within the profile HMMs, as well as transitions or ‘jumps’, between the different profile HMMs, at almost any position in the alignment. And positions of jumps between different subtypes define recombination breakpoints. The output report for each sequence comprises the predicted recombination and breakpoint intervals in text format, as well as a graphical representation of the predicted recombinant

fragments within the HIV-1 genome (Schultz, et al., 2009; Schultz, et al., 2007)

### 2.8.3 Subtype Classification Using Evolutionary Algorithms

Subtype Classification Using Evolutionary Algorithms (SCUEAL) was used to test for both intra and inter recombinants in the partial HIV-1 *pol* and IN sequences (http://www.datamonkey.org/dataupload_scueal.php). SCUEAL is a model-based phylogenetic method for subtyping HIV-1 sequences, mapping the location of breakpoints and assigning parental sequences in recombinant strains as well as computing confidence levels for the inferred quantities. SCUEAL is a completely automatic method which returns a predicted subtype, existing CRF or a recombinant form mapped in terms of a Bayesian framework, and every estimated quantity including the recombinant structure, the location of each breakpoint and the assignment of a parental/sister lineage is estimated with statistical confidence to allow an objective evaluation of how robust the estimates are, and the algorithm runs sufficiently quickly to permit the screening of thousands of sequences on a computer cluster. SCUEAL also screens a single sequence against a fixed reference alignment. SCUEAL runs on the datamonkey.org platform and accepts large reference sequence alignments which can be easily updated when new references, for example CRF, become available (Delport, Poon, Frost, Kosakovsky Pond, 2010; Kosakovsky Pond, et al., 2009).

## 2.9 Phylogenetic analysis

### 2.9.1 Multiple alignments and editing of sequences.

Multiple sequence alignments comprising the partial *gag*, *pol* and *env* sequences, and reference sequences derived from HIV-1 M group subtypes A-D, F-H, J-K and circulating recombinant forms obtained from the LANL database (http://www.hiv.lanl.gov/), was constructed using Clustal X v1.81. The isolate N.CM.95.YBF30.AJ006022 was used as an out-group in the alignments. The Clustal X multiple alignment algorithm involves calculating the distance matrix giving the divergence of each pair of sequence by aligning all pairs of sequences separately, using the resulting distance matrix to create

a guide using the neighbor joining method, and then aligning the sequences progressively according to the branching order in the guide tree (Thompson, Plewniak, Poch, 1999; Thompson, Gibson, Plewniak, Jeanmougin, Higgins, 1997). Clustal imposes penalties for gaps, mismatches, as well as for breaking secondary structure (Egan and Crandall, unpublished). The multiple sequence alignments were subsequently edited manually using BioEdit v 5.09 (Hall, 2005).

**Table 2.11.** HIV-1 subtypes A – K and CRF reference sequences from the Los Alamos National Laboratory HIV Sequence Database used in the construction of phylogenetic trees.

| Sequence Name | Accession Number | HIV-1 subtype/ group | Location | Year of isolation | Reference |
|---|---|---|---|---|---|
| Q23_17 | AF004885 | A1 | Kenya | 1994 | Neilson et al., 1999 |
| 97CDKTB48 | AF286238 | A2 | CD | 1997 | Gao et al., 2001 |
| HXB2_LAI_IIIB_BRU | K03455 | B | France | 1983 | Van Beveren et al., 1985 |
| 1058_11 | AY331295 | B | United States | 1998 | Bernardin et al., 2005 |
| 92BR025_d | U52953 | C | Brazil | 1992 | Gao et al., 1996. |
| 00BW0874.21 | AF443090 | C | Botswana | 2000 | Novitsky et al., 2002. |
| ETH2220 | U46016 | C | Ethiopia | 1986 | Salminen et al., 1996. |
| 95IN21068 | AF067155 | C | India | 1995 | Lole et al., 1999 |
| 93MW_965 | AY713413 | C | Malawi | 1993 | Brown et al., 2005 |
| 01TZA246 | AY253308 | C | Tanzania | 2001 | Arroyo et al., 2004 |
| 04ZASK146B1 | AY772699 | C | South Africa | 2004 | Rousseau et al., 2006 |
| 02ZM110C07 | AB254142 | C | Zambia | 2002 | Tatsumi et al.,unpublished |
| 01TZA280 | AY253311 | D | Tanzania | 2001 | Arroyo et al., 2004 |
| 93BR020_1 | AF005494 | F1 | Brazil | 1993 | Gao et al., 1998 |
| CM53657 | AF377956 | F2 | Cameroon | 1997 | Carr et al., 2001 |
| 92NG083 | U88826 | G | Nigeria | 1992 | Gao et al., 1998 |
| 90CR056 | AF005496 | H | CAR | 1990 | Gao et al., 1998 |
| SE7022 | AF082395 | J | Sweden | 1994 | Laukkanen et al., 1999 |
| 96CM_MP535 | AJ249239 | K | Cameroon | 1996 | Triques et al., 2000 |
| CM240 | U54771 | CRF01_AE | Thailand | 1990 | Carr et al., 1996 |
| 04BR142 | AY727527 | CRF31_BC | Brazil | 2004 | Sanabani et al., 2006 |
| 96TZ_BF071 | AF289549 | CRF10_CD | Tanzania | 1996 | Koulinska et al., 2001 |
| 00CMNYU1162 | EF087995 | CRF36_cpx | Cameroon | 2000 | Powell et al., 2007 |
| YBF30 | AJ006022 | N | Cameroon | 1995 | Simon et al., 1998 |

### 2.9.2 Neigbor Joining method using MEGA

The phylogenetic trees for the different HIV-1 genetic fragments were constructed using the Neighbor-Joining method implemented in the Molecular Evolutionary Genetic Analysis version 5 software (MEGA v5) (Tamura, Peterson, Peterson, Stecher, Nei, Kumar, 2011).

The reference sequences were derived from reference sequences from the LANL database (http://www.hiv.lanl.gov/) (**Table 2.11**).

After the alignments were performed and manually edited, multiple alignment files were imported into MEGA v 5 where the alignment files (.pir) were converted into a MEGA format (.meg). The MEGA files (.meg) were then opened in MEGA and NJ-trees (Saitou and Nei, 1987) were constructed with the use of the Kimura 2-parameter (Kimura, 1980). Bootstrap analysis was also performed on the trees to confer statistical significance, with a total of a 1,000 bootstrap replicates for each dataset (Felsenstein, 1985).

### 2.9.3 Maximum Likelihood method using MEGA

The phylogenetic trees for the different HIV-1 genetic fragments were constructed using the Maximum Likelihood method implemented in the Molecular Evolutionary Genetic Analysis version 5 software (MEGA v5) (Tamura, Peterson, Peterson, Stecher, Nei, Kumar, 2011; Felsenstein, 1981). A General Time Reversible plus discrete Gamma distribution was used to model evolutionary rate differences among sites (GTR+Γ) (Tavaré, 1986; Yang, 1994).

## 2.10 Transmitted drug resistance using CPR

The calibrated population resistance (CPR) tool is a web-accessible program for performing standardized genotypic estimation of transmitted HIV-1 drug resistance. The program is linked to the Stanford HIV drug resistance database and can additionally perform viral genotyping and algorithmic estimation of resistance to specific antiretroviral drugs.The CPR program accepts FASTA-formatted HIV-1 PR and RT sequence data. A profile alignment of the submitted sequence set is created by aligning each nucleotide sequence to a polypeptide reference sequence for the region of the HIV-1 genome encoding PR and RT, which by default is a subtype B consensus sequence available from http://hivdb.stanford.edu/ .

A list of DRMs, by default the most recent version of the SDRM list, is used to compute the prevalence of resistance to each of the three main classes of antiretroviral drug: protease inhibitors (PIs), nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs). The prevalence of transmitted HIVDR to each drug class is

estimated as the number of sequences containing any DRM specific to that drug class relative to the number of times the target gene is represented in the alignment. Analysis generates a report that summarizes the input data set in terms of drug resistance, genetic diversity, and sequence quality. The CPR report includes a graphical overview of DRMs and resistance-associated mutations present in the input dataset and a plot showing coverage across the target region (Gifford et al., 2009).

CPR was used to test for transmitted drug resistance using partial HIV-1 *pol* sequences http://cpr.stanford.edu/cpr.cgi. Mutations, deletions, and insertions, defined as changes relative to the reference sequence, are recorded for each submitted sequence. The prevalence of individual mutations is calculated by dividing mutation frequency by the number of valid codons at the corresponding position in the alignment (Gifford et al., 2009).

### 2.10.1 HIV drug resistance mutations using HIVdb

HIV-1 PR and RT antiretroviral drug resistance mutations were determined using the Stanford University HIV Drug Resitance Database (HIVdb), http://www.hivdb.stanford.edu/DR/asi/releaseNotes/index.html.

The database links sequence changes in the molecular targets of HIV-1 therapy to other forms of data including treatment history and phenotypic, or drug susceptibility, data. HIVdb accepts user-submitted HIV-1 *pol* sequences and returns inferred levels of resistance to 20 FDA-approved ARV drugs including 8 PIs, 7 NRTIs, 4 NNRTIs, and 1 INI (Rhee, Gonzales, Kantor, Betts, Ravela, Shafer, 2003).

# CHAPTER THREE:

# RESULTS

# CHAPTER THREE

# RESULTS

## 3.1. Patient cohort

The PIME study involved 51 plasma samples that were collected from a patient cohort from Bushbuckridge, Mpumalanga, between February and July 2009. The CD4+ T lymphocyte count of the samples varied between 105 cells and 1263 cells (**Table 3.1**). None of the patients were on antiretrovirals.

## 3. 2. PCR amplification of partial *gag*, *pol* and *env* fragments.

PCR amplification was successful for most of the Bushbuckridge, Mpumalanga samples, with 74.5% of the partial *gag* p24 gene, 54.9% of the partial IN p32 gene, and 83.3% of the partial *env* gp41 showing positive bands in an agarose gel after electrophoresis. However, the PCR amplification of the partial *pol* p66/p51 gene was considerably less successful at 29%. The PCR amplification of only one sample, that is sample 0040 or 2% of the samples, was negative for all sub-genomic regions (**Table 3.1**). Representative gels of the four different PCR assays are shown is **Figure 3.1** to **Figure 3.4.**

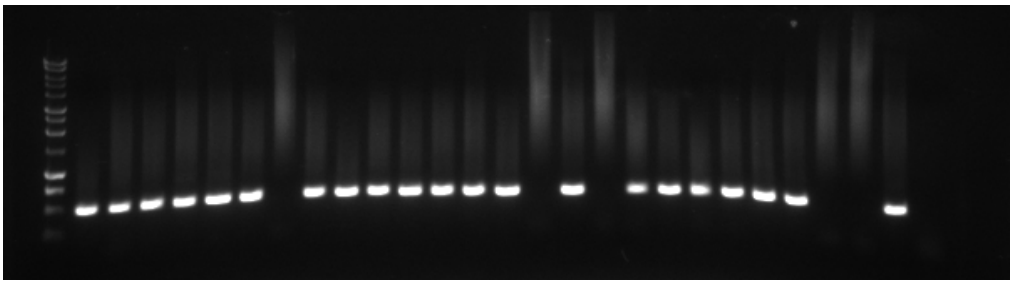


**Figure 3.1.** A 0.8% agarose gel analysis of the PCR amplification of the 484 bp fragment of the *gag* gene comprising a sequence that encodes the HIV-1 CA protein p24, and which is derived from isolates of the cohort in Mpumalanga Province. Lane 1: 1kb DNA Ladder.

**Table 3.1.** Patient samples and PCR results (POS, Positive; –, Negative; ND, Not done).

| Patient Number | CD4+ cell count | Date of specimen collection | gag p24 PCR | Pol p66/p51 PCR | IN p32 PCR | env gp41 PCR |
|---|---|---|---|---|---|---|
| 0005A | 466 | 23 April 2009 | POS | – | POS | POS |
| 0018A | ND | ND | POS | – | POS | POS |
| 0022A | 731 | 24 April 2009 | POS | POS | POS | POS |
| 0038 | 680 | 16 March 2009 | POS | POS | POS | POS |
| 0039a | 307 | 17 March 2009 | POS | POS | POS | POS |
| 0040A | 506 | 23 March 2009 | POS | – | – | POS |
| 0040 | ND | 23 March 2009 | – | – | – | – |
| 0042A | 511 | 19 February 2009 | POS | POS | POS | POS |
| 0043A | ND | ND | – | – | – | POS |
| 0055A | ND | ND | POS | POS | POS | POS |
| 0064A | 105 | 16 March 2009 | POS | – | – | POS |
| 0066A | 583 | 17 March 2009 | POS | – | POS | POS |
| 0073A | 366 | 2 April 2009 | POS | – | – | POS |
| 0073 | ND | ND | POS | – | POS | POS |
| 0081A | 437 | 8 April 2009 | POS | – | – | POS |
| 0085 | 154 | 14 April 2009 | – | – | – | POS |
| 0092A | 137 | 16 April 2009 | POS | – | POS | POS |
| 0097A | 243 | 17 April 2009 | – | – | POS | POS |
| 0098A | 261 | 20 April 2009 | POS | POS | – | POS |
| 0101A | 790 | 20 April 2009 | POS | POS | – | POS |
| 0103 | 191 | 20 April 2009 | POS | – | – | POS |
| 0116A | 150 | 29 July 2009 | POS | – | POS | POS |
| 0119A | 403 | 29 April 2009 | POS | POS | POS | POS |
| 0122A | 262 | 30 April 2009 | POS | POS | – | POS |
| 0123A | 335 | 4 May 2009 | – | – | – | POS |
| 0130A | 387 | 6 May 2009 | – | – | POS | POS |
| 0132A | 489 | 7 May 2009 | POS | – | – | POS |
| 0134A | 792 | 11 May 2009 | POS | POS | – | POS |
| 0135A | 1263 | 11 May 2009 | POS | – | – | POS |
| 0136A | 1192 | 11 May 2009 | POS | POS | POS | POS |
| 0143A | 353 | 25 May 2009 | POS | – | POS | POS |
| 0147A | 785 | 27 May 2009 | POS | POS | – | POS |
| 0152A | 691 | 3 June 2009 | POS | – | – | POS |
| 0155A | ND | ND | – | – | – | POS |
| 0165A | 198 | 22 June 2009 | POS | – | – | POS |
| 0166A | ND | ND | POS | – | – | POS |
| 0173A | 367 | 29 June 2009 | POS | – | POS | POS |
| 0185A | 311 | 2 July 2009 | – | – | POS | POS |
| 0189 | ND | 20 July 2009 | | | | |
| 0189A | 560 | 6 July 2009 | POS | POS | POS | POS |
| 0190A | 229 | 6 July 2009 | POS | POS | – | – |
| 0192A | 522 | 6 July 2009 | – | G | POS | POS |
| 0193A | 313 | 6 July 2009 | – | – | POS | POS |
| 0193 | ND | 6 July 2009 | POS | – | – | POS |
| 0198A | 217 | 7 July 2009 | – | POS | POS | POS |
| 0199A | 219 | 7 July 2009 | POS | – | – | POS |
| 0203A | 401 | 8 July 2009 | – | POS | – | – |
| 0204A | 349 | 28 July 2009 | POS | – | – | POS |
| 0206A | 314 | 8 July 2009 | – | – | POS | – |
| 0207A | 733 | 8 July 2009 | POS | POS | POS | – |
| 0211 | 312 | 28 July 2009 | POS | – | – | POS |
| 0215A | 726 | 28 July 2009 | POS | – | – | POS |

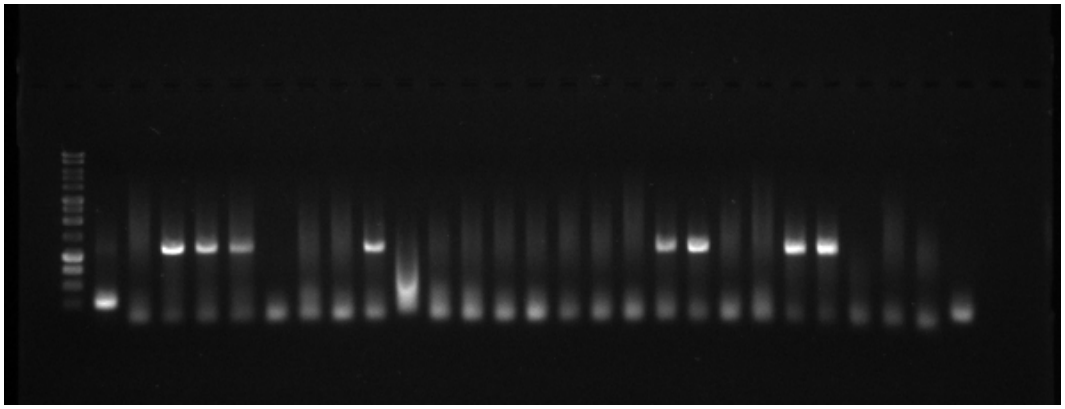


**Figure 3.2.** Agarose gel analysis of the PCR amplification of the 1085 bp fragment of the *pol* gene comprising a sequence that encodes the HIV-1 PR and RT proteins, p10 and p61/p51 respectively. Lane 1: 1kb DNA Ladder; lanes 2-26, patient samples. Lanes 8, 16, 18 and 25 were negative.

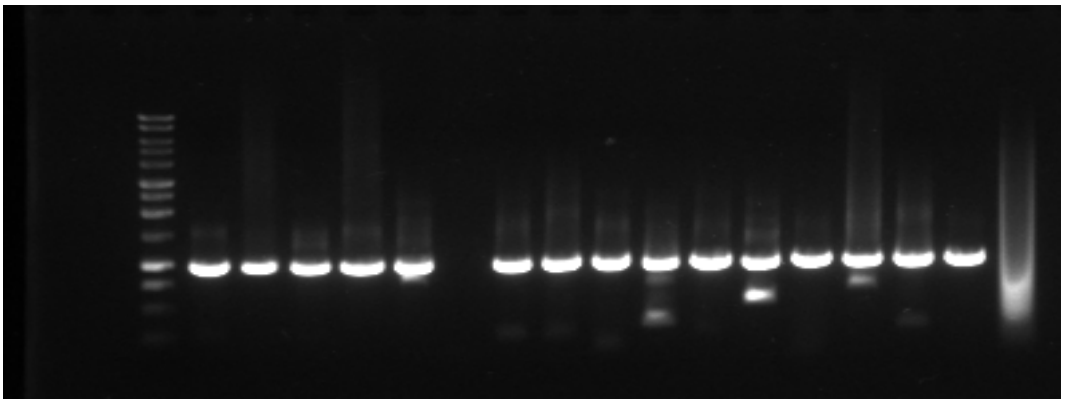


**Figure 3.3.** Agarose gel analysis of the PCR amplification of the 944 bp fragment of the *pol* gene comprising a sequence that encodes the HIV-1 IN protein p32. Lane 1: 1kb DNA Ladder, lanes 2-18, patient samples. Lanes 8, 16, 18, 25 and 26 were negative.

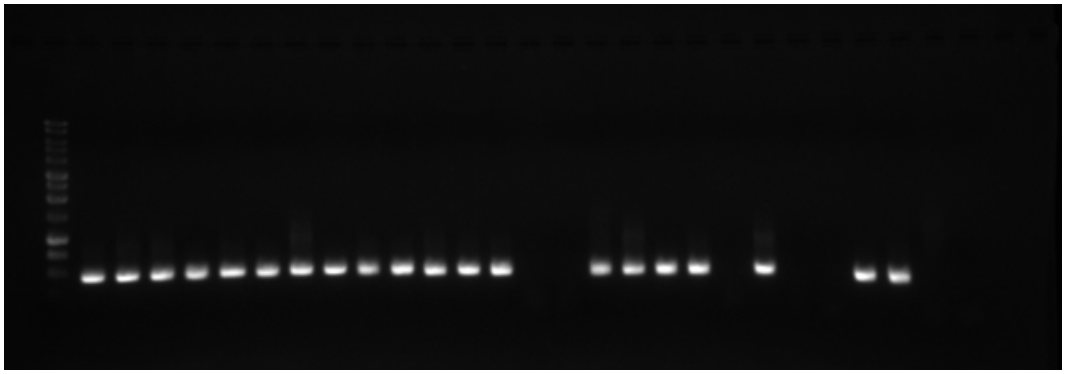


**Figure 3.4.** Agarose gel analysis of the PCR amplification of the 438 bp fragment of the *env* gene comprising a sequence that encodes the HIV-1 TM glycoprotein gp41. Lane 1: 1kb DNA Ladder, lanes 2-26, patient samples. Lanes 8, 16, 18, 25 and 26 were negative.

### 3.3 Sequencing data and quality control.

Forty of the *gag* p24, 15 of the *pol*, RT/PR, 22 of the *pol* IN and 44 of the *env* gp41 amplicons were successfully sequenced.

When the LANL QC tool indicated stop codons, the sequences were reread and verified. Hypermutation was detected in two Integrase sequences: 0130A and 0207A. (Detailed information is indicated in the **Appendix**).

Subtypes were assigned using the RIP tool (**Table 3.2**). All *gag* sequences were assigned HIV-1 subtype C, except for: 0042A, 0189A, 0198A and 203. All *pol* sequences were assigned HIV-1 subtype C, except for: 00132A, 0192A, 0198A and 204A. All *env* sequences were assigned HIV-1 subtype C, except for: 00143A and 0043A.

### 3.4 Subtype analysis using jpHMM and REGA.

Detailed information is indicated in the **Appendix**. REGA and jpHMM online tools were used to assign subtypes to 42 *gag*, 14 *pol* p10 and p66/p51, 21 *pol* p32 and 43 *env* gp41 sequences. The jpHMM subtyping tool assigned 38 *gag* sequences, 13 *pol* sequences, 19 *pol* p32 sequences to subtype C, while the REGA subtyping tool assigned all sequences to subtype C except a single sequence.

REGA detected a C, F1 intersubtype recombinant in the 0042A *gag* genetic fragment, which was discrepant from the results found in the same *gag* genetic fragment by the jpHMM subtyping tool, which is an HIV-1 subtype F2 without any recombinant breakpoints. Significantly, the jpHMM subtyping tool and the RIP from the LANL QC detected recombinants in the same genetic fragments from the same isolates. Barring the fact that the recombinants detected by the the jpHMM subtyping tool and the RIP are discrepant (see **Table 3.2** and **Table 3.3**) this is a remarkable instance of reproducibility by these subtyping and recombinant detection tools.

**Table 3.2.** Subtyping of the partial *gag* p24, *pol* p66/p51 and p32 and *env* gp41 regions using RIP (–, genomic regions not subtyped).

| Sample | *gag* p24 | *pol* PR/RT | *pol* p32 | *env* gp41 |
|---|---|---|---|---|
| 0005A | C | – | C | C |
| 0018A | – | – | – | C |
| 0022A | – | – | – | C |
| 0038A | C | – | – | – |
| 0039A | C | – | – | C |
| 0040A | C | – | – | C |
| 0042A | B, C, F2, A2 | C | C | C |
| 0043A | – | – | – | B |
| 0055A | – | – | – | C |
| 0064A | C | C | C | C |
| 0066A | C | – | C | C |
| 0073 | – | – | C | C |
| 0073A | C | – | – | C |
| 0081A | C | – | C | C |
| 0085A | C | C | C | – |
| 0092A | C | – | – | C |
| 0097A | C | – | – | C |
| 0098A | C | – | C | C |
| 0101A | C | – | C | C |
| 0103A | C | – | – | C |
| 0116A | – | – | – | – |
| 0119A | C | – | – | – |
| 0122 | – | – | – | C |
| 0122A | C | – | – | C |
| 0123A | C | – | – | C |
| 0130A | C | C | C | C |
| 0132A | C | C, F1 | – | C |
| 0134A | C | – | – | C |
| 0135A | C | – | – | – |
| 0136A | C | C | C | – |
| 0143A | C | C | C | Indeterminate |
| 0143A_44 | – | C | – | – |
| 0147A | C | – | – | C |
| 0152A | C | – | – | C |
| 0155A | C | – | – | C |
| 0165A | C | – | – | C |
| 0166A | C | – | – | C |
| 0173A | C | C | C | C |
| 0185A | C | – | C | C |
| 0189A | A1, A2, B, C | C | C | C |
| 0190A | C | – | – | C |
| 0192A | C | C | C, 01_AE | C |
| 0193A | C | – | C | C |
| 0193 | C | – | – | C |
| 0198A | Indeterminate | – | C | C |
| 0199A | – | – | – | C |
| 0203A | C, D, A2, A1, AE, F2 | – | – | C |
| 0204A | C | Indeterminate | – | C |
| 0206A | C | C | C | C |
| 0207A | C | C | C | C |
| 0211A | C | – | – | C |
| 0215A | C | – | – | – |

**Table 3.3.** Subtyping of the partial *gag* p24, *pol* p66/p51 and p32 and *env* gp41 regions using jpHMM and REGA online tools. (–, genomic regions not subtyped)

| Sample | *gag* p24 | | *pol* p66/p51 | | *pol* p32 | | *env* gp41 | |
|---|---|---|---|---|---|---|---|---|
| | jpHMM | REGA | jpHMM | REGA | jpHMM | REGA | jpHMM | REGA |
| 0005A | C | C | – | – | C | C | C | C |
| 0018A | – | – | – | – | – | – | C | C |
| 0022A | – | – | – | – | – | – | C | C |
| 0038A | C | C | – | – | – | – | – | – |
| 0039A | C | C | – | – | – | – | C | C |
| 0040A | C | C | – | – | C | C | – | – |
| 0040 | – | G | C | C | – | – | C | C |
| 0042A | F2 | C, F1 | C | C | C | C | C | C |
| 0043A | – | – | – | – | – | – | B | B |
| 0055A | – | – | – | – | – | – | C | C |
| 0064A | C | C | C | C | – | C | C | C |
| 0066 | | C | – | – | C | | | |
| 0066A | C | C | – | – | C | C | C | C |
| 0073 | C | | – | – | – | – | C | C |
| 0073A | C | C | – | – | – | – | C | C |
| 0081A | C | C | – | – | C | C | C | C |
| 0085 | C | C | C | C | C | – | – | – |
| 0092A | C | C | – | – | – | – | C | C |
| 0097A | C | C | – | – | – | – | C | C |
| 0098A | C | C | – | – | C | C | C | C |
| 0101A | C | C | – | – | C | C | C | C |
| 0103A | C | C | – | – | – | – | C | C |
| 0116A | – | – | – | – | – | – | – | – |
| 0119A | C | | – | – | – | – | – | – |
| 0122A | B | C | – | C | – | – | C | C |
| 0123A | C | C | – | – | – | – | C | C |
| 0130A | C | C | C | C | C | C | C | C |
| 0132A | C | C | B, C | C | – | – | C | C |
| 0134A | C | C | – | – | – | – | C | C |
| 0135A | C | C | – | – | – | – | – | – |
| 0136A | – | – | C | C | C | C | – | – |
| 0143A | C | C | C | C | C | C | – | C |
| 0147A | C | C | – | – | – | – | C | C |
| 0152A | C | C | – | – | – | – | C | C |
| 0155A | – | – | – | – | – | – | C | C |
| 0165A | C | C | – | – | – | C | C | C |
| 0173A | C | C | C | C | C | C | C | C |
| 0185A | C | C | – | – | – | C | C | C |
| 0189A | G | C | C | C | C | C | C | C |
| 0190A | C | C | – | – | – | – | C | C |
| 0192A | C | C | C | C | B, C, H | C | C | C |
| 0193A | – | – | – | – | C, K | C | C | C |
| 0193 | C | C | – | – | – | C | – | C |
| 0198A | – | | – | – | C | C | C | C |
| 0199A | C | C | – | – | – | – | C | C |
| 0203A | F2 | C | – | – | – | – | C | C |
| 0204A | C | C | – | – | – | – | C | C |
| 0206A | C | C | C | C | C | C | C | C |
| 0207A | C | C | C | C | C | C | C | C |
| 0211A | C | C | – | – | – | – | C | C |
| 0215A | C | C | – | – | – | – | – | – |

### 3.5 Subtype analysis of partial *pol* fragments using SCUEAL.

The SCUEAL subtyping of the *pol* PR/RT and IN gene fragments revealed that, unlike the REGA and jpHMM subtyping tools, the PR/RT and IN sequences were HIV-1 subtype C (**Table 3.4**). Detailed information is listed in the **APPENDIX**.

**Table 3.4.** jpHMM, REGA and SCUEAL subtyping of the partial *pol* p66/p51 and *pol* IN p32 gene fragments. (–, genomic regions not subtyped)

| Sample | *pol* p66/p51 | | | *pol* p32 | | |
|---|---|---|---|---|---|---|
| | jpHMM | REGA | SCUEAL | jpHMM | REGA | SCUEAL |
| 0005A | – | – | C | C | C | C |
| 0040 | – | – | C | C | C | C |
| 0042A | C | C | C | C | C | C |
| 0064A | C | C | C | – | C | C |
| 0066A | – | – | – | C | C | C |
| 0081A | – | – | – | C | C | C |
| 0085 | C | C | C | C | – | C |
| 0098A | – | – | – | C | C | C |
| 0101A | – | – | – | C | C | C |
| 0122A | – | C | | – | – | |
| 0130A | C | C | C | C | C | C |
| 0132A | B, C | C | C | – | – | – |
| 0136A | C | C | C | C | C | C |
| 0143A | C | C | C | C | C | C |
| 0165A | – | – | – | – | C | |
| 0173A | C | C | C | C | C | C |
| 0185A | – | – | – | – | C | C |
| 0189A | C | C | C | C | C | C |
| 0190A | – | – | – | – | – | – |
| 0192A | C | C | C | B, C, H | C | C |
| 0193A | – | – | – | C, K | C | C |
| 0198A | – | – | – | C | C | C |
| 0206A | C | C | C | C | C | C |
| 0207A | C | C | C | C | C | C |

### 3.6 Possible recombinant sequences detected using online tools.

The online tools indicated some possible recombinant sequences (**Table 3.2** to **Table 3.4**; **Figure 3.5** to **Figure 3.10**), but the results were not always supported (**Appendix**).

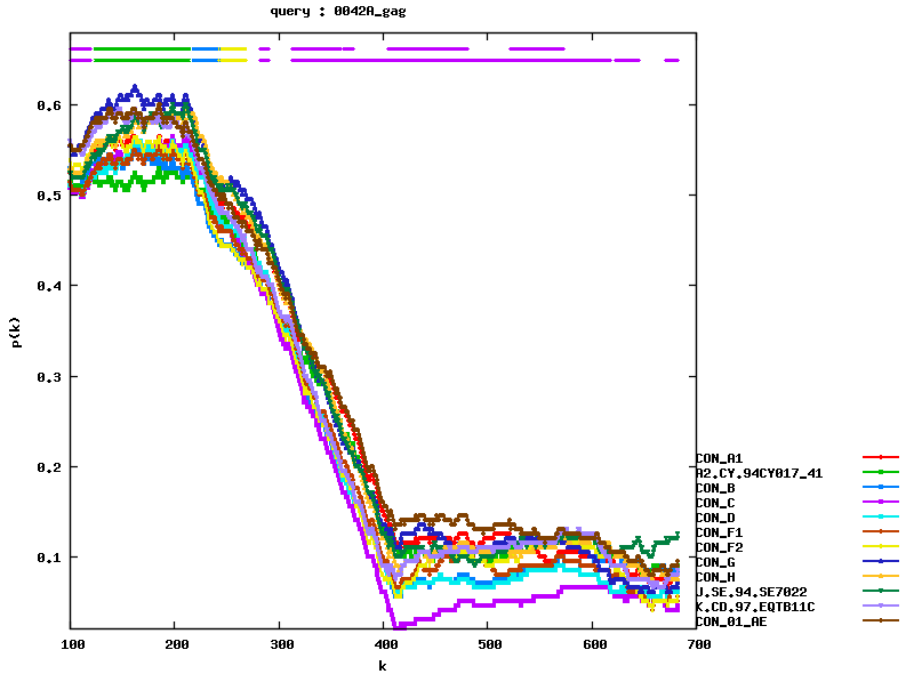
The LANL QC graphs show similarity plots of the query sequence against various background sequences in an alignment in a sliding window. The the x-axis (k) represents the query sequence position at the center of the moving window and the y-axis s(k) shows the similarity between that window of sequence and each of the background sequences. The similarity plot for 0042A_gag shows consensus G is the sequence with the highest similarity in the background alignment (**Figure 3.5**). The two bars across the top of the graph represent the best match, lower bar, and the significance of the match, upper bar. The best match sequence is the background sequence with the highest similarity to the query, which in this case is consensus C.

In **Figures 3.5 – 3.10**, the REGA and jpHMM graphs show maps of the HIV genome with the different open reading frames and the relevant genetic fragments with the designated HIV-1 subtypes indicated through colour coding. The REGA graphs show bootstrap support for the subtype, and the jpHMM graphs show the positions of the query sequences with respect to their start, end and possible breakpoints with reference to the HIV-1 HXB2 reference sequence.

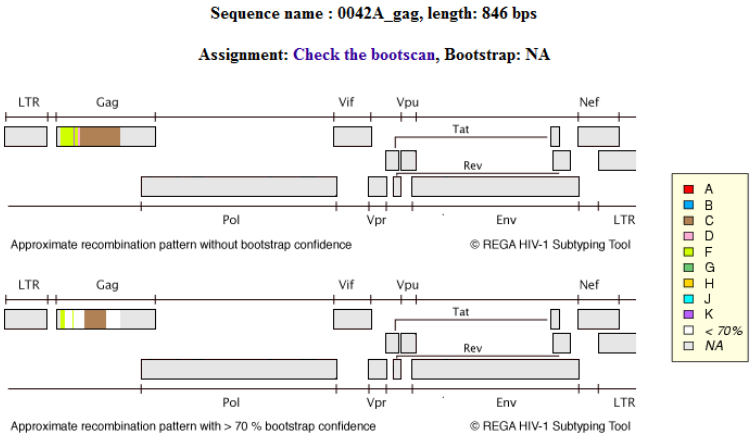
The analysis of the *gag*, *pol* p10 and p66/p51, *pol* p32 and *env* gp41 genetic fragments of the isolates from Bushbuckridge, Mpumalanga using online quality control and subtyping tools has revealed that, while most of these fragments are clearly HIV-1 subtype C, a significant number may be inter-subtype recombinants. Different genetic fragments from 0042A_gag, 0189A_gag, 0203A_gag 0132A_pol, 0192A_IN and 0193A_IN consistently exhibit recombination breakpoints across all online platforms, that is, the LANL QC tool, jpHMM, REGA. However, SCUEAL seems to have failed to detect any recombination breakpoints in the genetic fragments from these isolates, including intra-subtype recombination, in the sequences of the

genetic fragments derived from these isolates. Four *pol* RT/PR partial sequences, 0042A, 0143A, 0173A, 0207A, were intra-subtype recombinants with breakpoints ranging from 1 to 4. And six IN partial sequences, 0040, 0143, 0173A, 0206A, 0081A and 0098, were intra-subtype recombinants with breakpoints ranging from 1 to 3. The *gag* sub-genomic region of the 0042A_gag, 0189A_gag, 0203A_gag isolates, the *pol* subgenomic region of the 0132A_pol isolate, and the IN sub-genomic region of the 0192A_IN 0193A_IN isolate thus, by virtue of results from most subtyping online tools used in this study.

**A**




**B**




**C**

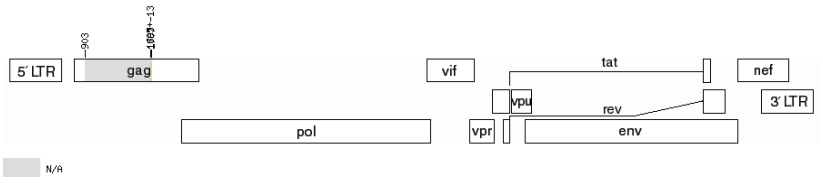
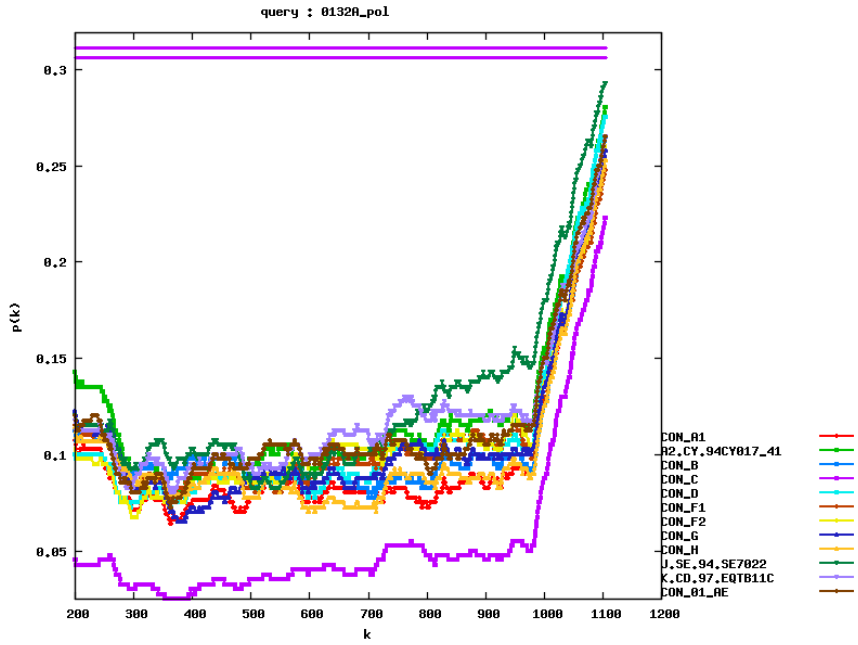
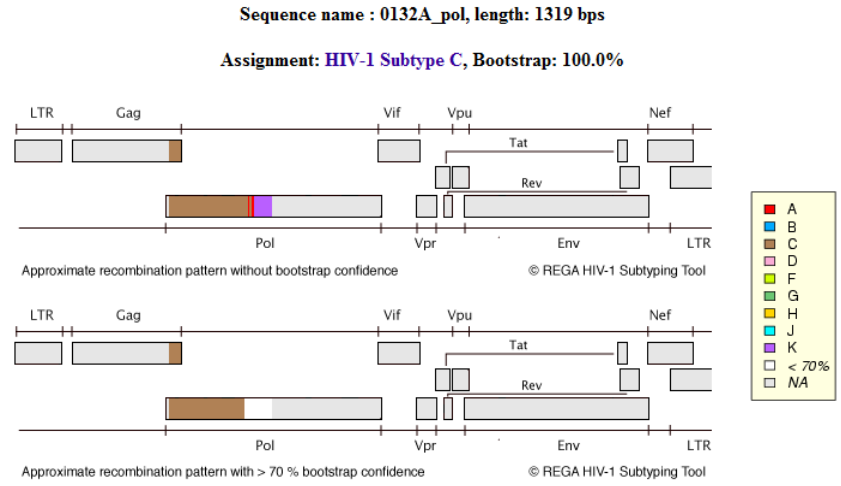



**Figure 3.5.** Possible recombinant 0042A_gag. **A:** LANL QC RIP detected a B, C, F2, A2 mosaic. **B:** REGA found a C, F1 recombinant. **C:** jpHMM found a F2 subtype.

**A**




**B**




**C**

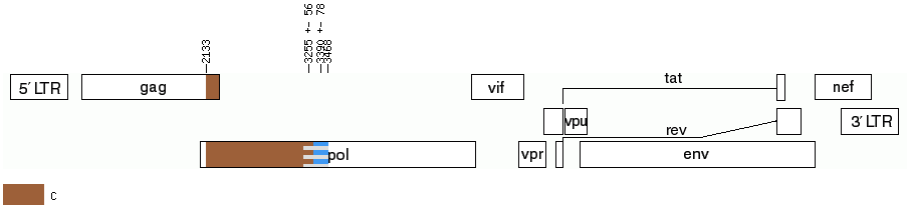
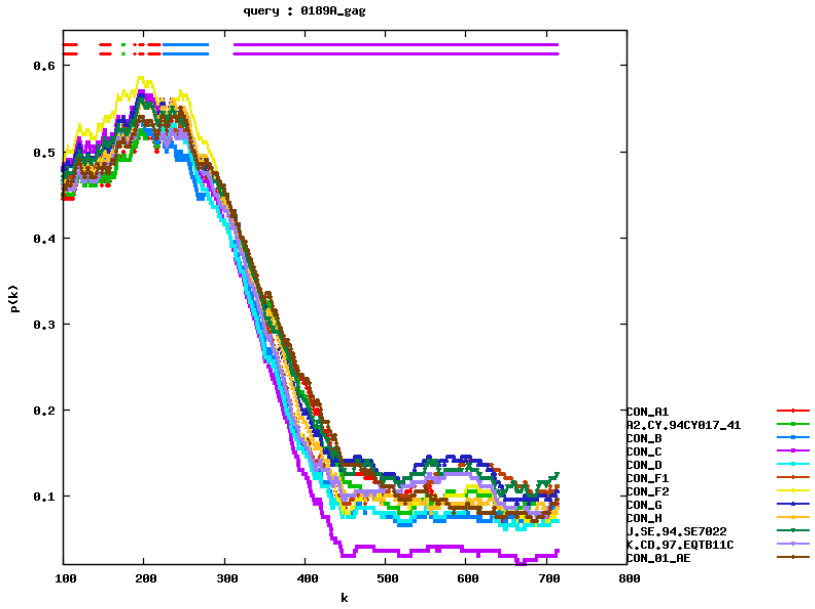
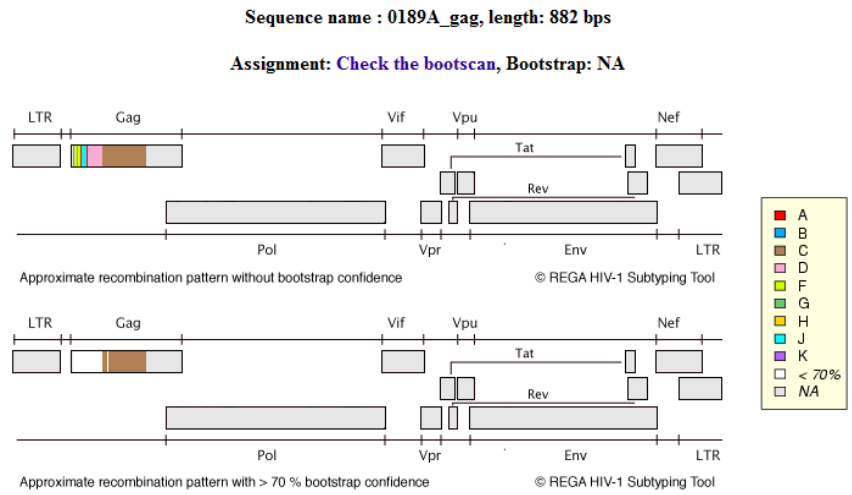



**Figure 3.6**. Possible recombinant 0132A_pol**. A:** LANL QC RIP found a C, F1 recombinant. **B:** REGA found a C subtype. **C:** jpHMM detected a B, C recombinant.

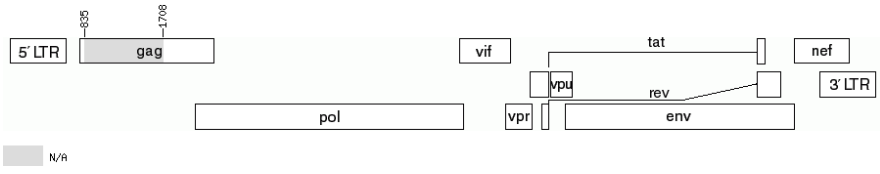
**A**




**B**



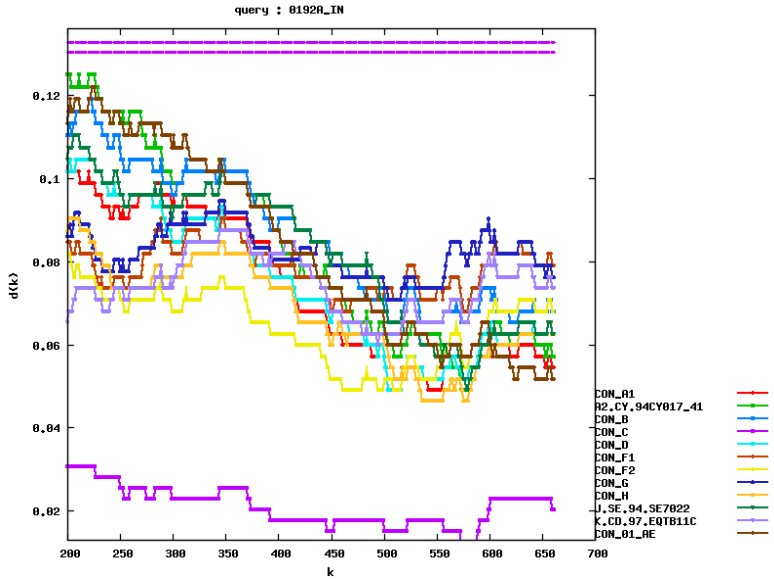

**C**



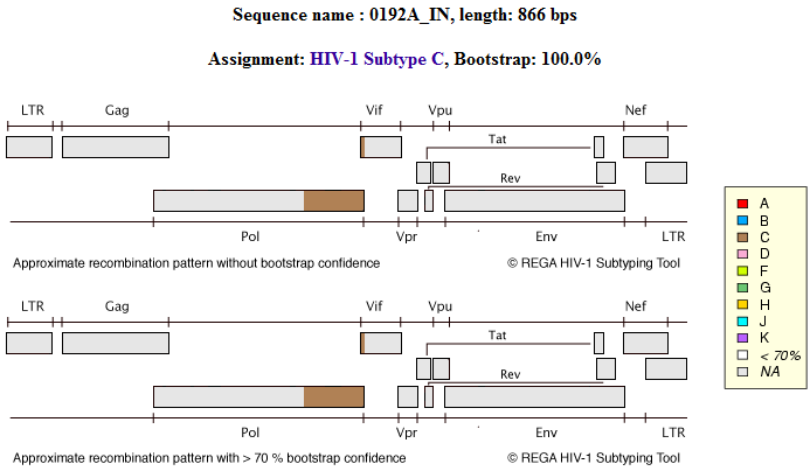

**G**

**Figure 3.7**. Possible recombinant 0189A_gag. **A:** LANL QC detected an A1, A2, B and C mosaic. **B:** REGA found a C sutype. **C:** jpHMM found a G subtype.

**A**




**B**




**C**

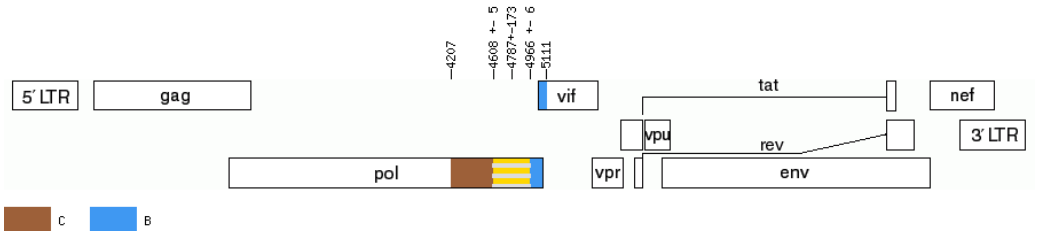
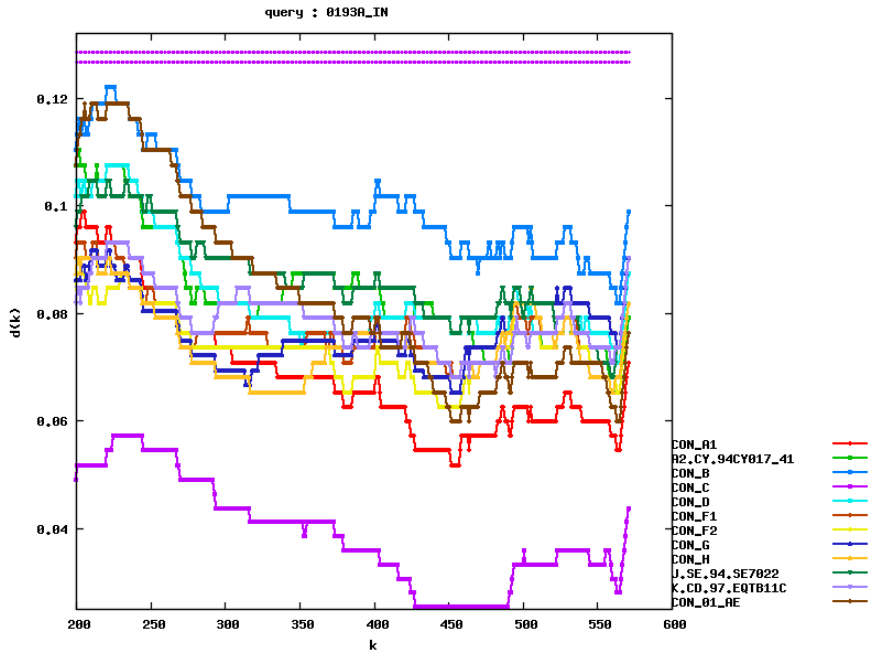
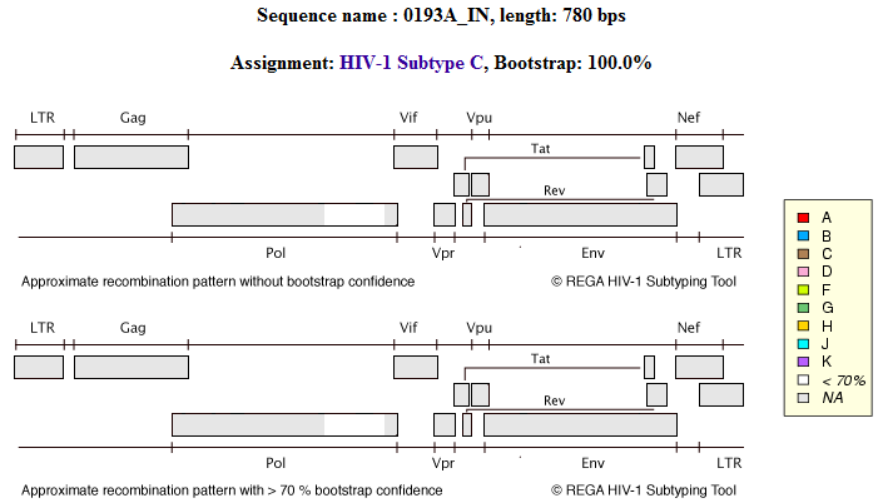



**Figure 3.8**. Possible recombinant 0192_IN. **A:** LANL QC detected a C, 01_AE recombinant. **B:** REGA determined a C subtype. **C:** jpHMM found a B, C, H mosaic.

**A**




**B**




**C**

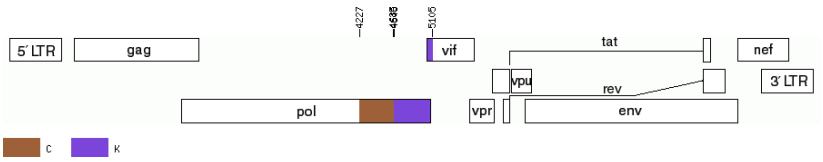
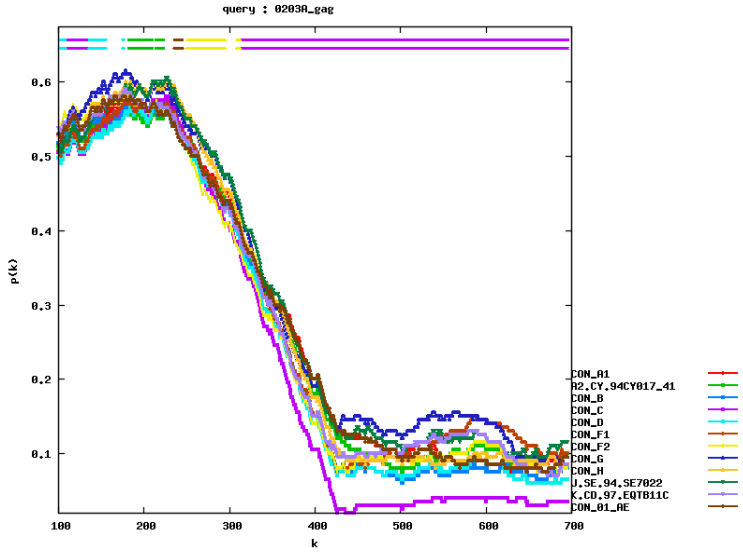
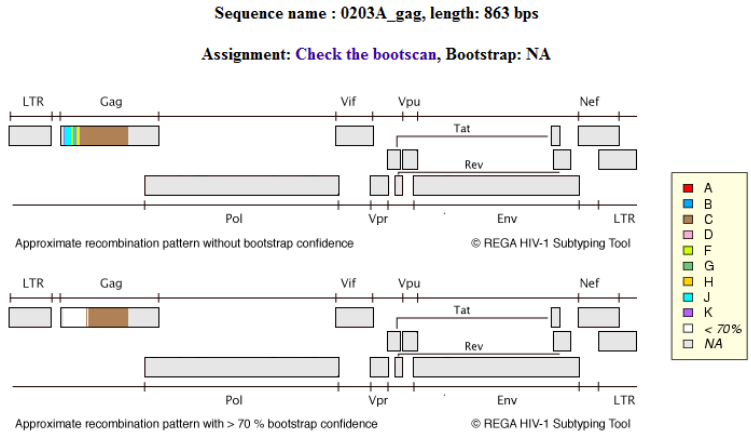



**Figure 3.9** Possible recombinant 0193_IN. **A:** LANL QC determined a C subtype. **B:** REGA found a C subtype. **C:** jpHMM detected a C, K recombinant.

**A**




**B**




**C**

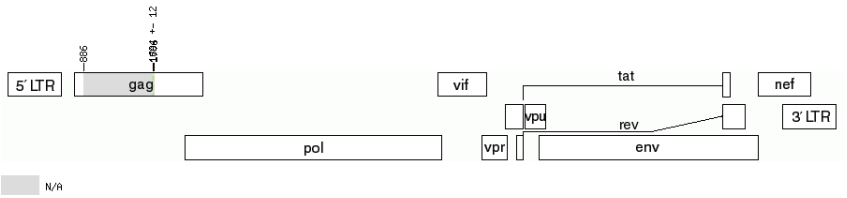



**Figure 3.10.** Possible Recombinant 0203A_gag. **A:** LANL QC found a C, D, A2, A1, 01_AE, F2 mosaic. **B:** REGA found a C subtype. **C:** jpHMM found an F2 subtype.

## 3.7 Phylogenetc analysis using MEGA

### 3.7.1 Phylogenetic trees of *gag* p24

The phylogenetic tree for the *gag* p24 gene comprised 57 sequences, and was aligned using 24 references sequences from the Los Alamos National Laboratory HIV Sequence Database. The evolutionary history was inferred using the Neighbor-Joining and the Maximum Lkelihood methods. All sequences clustered with HIV-1 subtype C reference sequences **(Figures 3.11 and 3.12)**.

### 3.7.2 Phylogenetic trees of *pol* p10 and p66/p51

The phylogenetic tree for the *pol* p66/p51 gene comprised 37 sequences including 24 reference sequences, of which 8 were subtype C sequences. The evolutionary history was inferred using the Neighbor-Joining and the Maximum Likelihood methods. All sequences clustered with HIV-1 subtype C reference sequences **(Figures 3.13, 3.14 and 3.15)**.

### 3.7.3 Phylogenetic trees of *pol* p32

The phylogenetic tree for the partial IN p32 contained 35 sequences, 25 references including 8 subtype C sequences. The evolutionary history was inferred using the Neighbor-Joining and the Maximum Lkelihood methods. All sequences clustered with HIV-1 subtype C reference sequences **(Figures 3.16 and 3.17)**.

### 3.7.4 Phylogenetic trees of *env* gp41

The phylogenetic tree for the *env* gp41 contained 55 sequences, 24 reference sequences including 8 subtype C sequences. The evolutionary history was inferred using the Neighbor-Joining and the Maximum Lkelihood methods. All sequences clustered with HIV-1 subtype C reference sequences, except for the 0043_gp sequence that clustered with HIV-1 subtype B sequences. **(Figures 3.18 and 3.19)**.

The phylogenetic analysis of the sequences of the *gag* p24, *pol* p66/p51 IN p32 and *env* gp41 sub-genomic regions of HIV-1 isolates from Bushbuckridge

shows that they are all, with the exception of the 0043A_gp *env* gp41 sequence, HIV-1 subtype C. While the 0132A_pol sequence doesn't bifurcate in the Neighbor-Joining trees, it bifurcates in the Maximum Likelihood tree. In fact 0132A_pol seems almost identical to the 0207A_pol sequence in the Neighbor-Joining trees, which might result from from sample mix-up or contamination for that region. None of the recombinant sequences detected through the LANL QC, jpHMM and REGA subtyping online tools clustered with reference sequences other than subtype C during phylogenetic analysis.

None of the sub-genomic regions supported a monophyletic South African lineage. The subtype C sequences from South Africa and southern African countries, as well as the sequence from India, all intermingle. Only sequences from Brazil (BR) formed country-specific cluster or monophyletic group. The *env* gp41 tree has a very distinctive branching structure with all the lineages splitting occurring only a short distance from the root so that the tree appears very 'bottom heavy'. It is possible that such a pattern reflects a very rapid diversification (high birth rate) of HIV in the early stages of the AIDS epidemic when there were a large number of susceptible hosts and established networks of transmission followed by a slowing- down in the rate of transmission (low birth rate) due to a decline in the number of susceptible hosts and the long asymptomatic period of HIV infection (Holmes and Garnett, 1994).

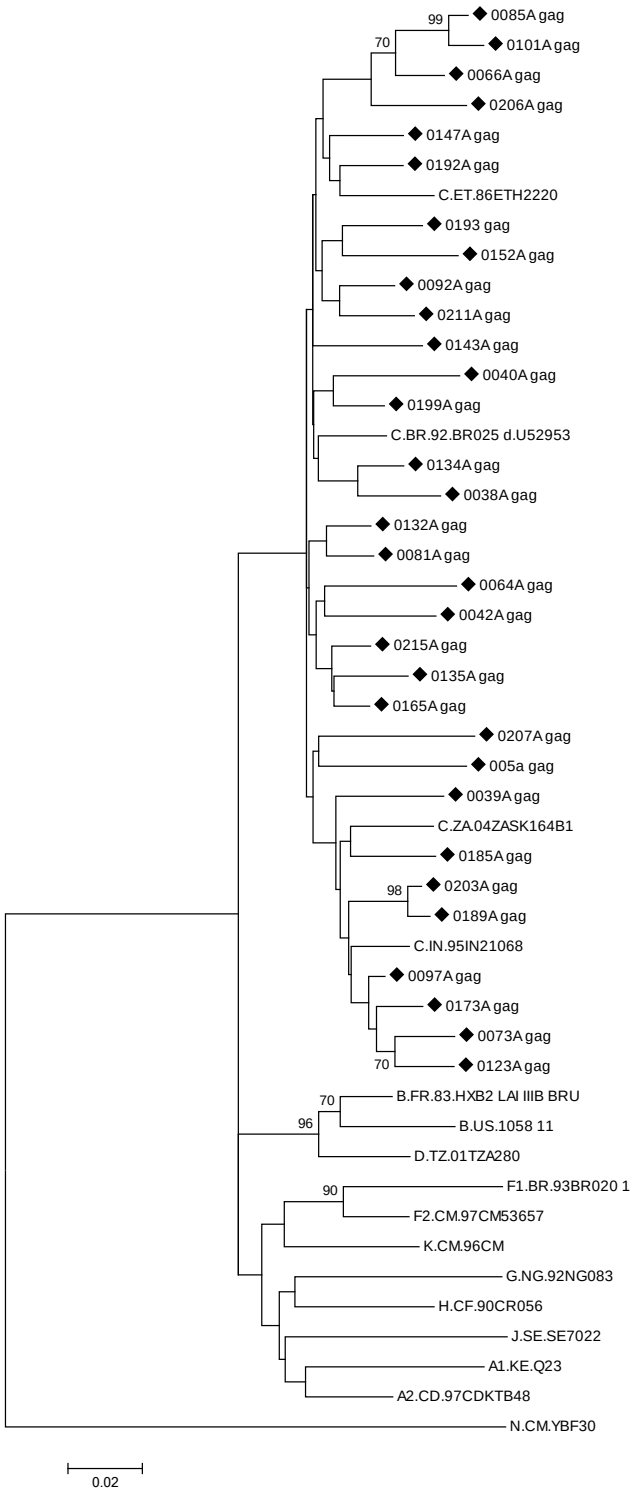



**Figure 3.11** Phylogenetic tree of the HIV-1 *gag* p24 genomic region. The Neighbor-Joining tree was constructed using the Kimura 2-parameter evolutionary model implemented in MEGA version 5. The bootstrap values of 1000 replicates above 70% are indicated next to the nodes and the scale bar at the bottom represents the number of base substitutions per site. The diamond indicates study samples.

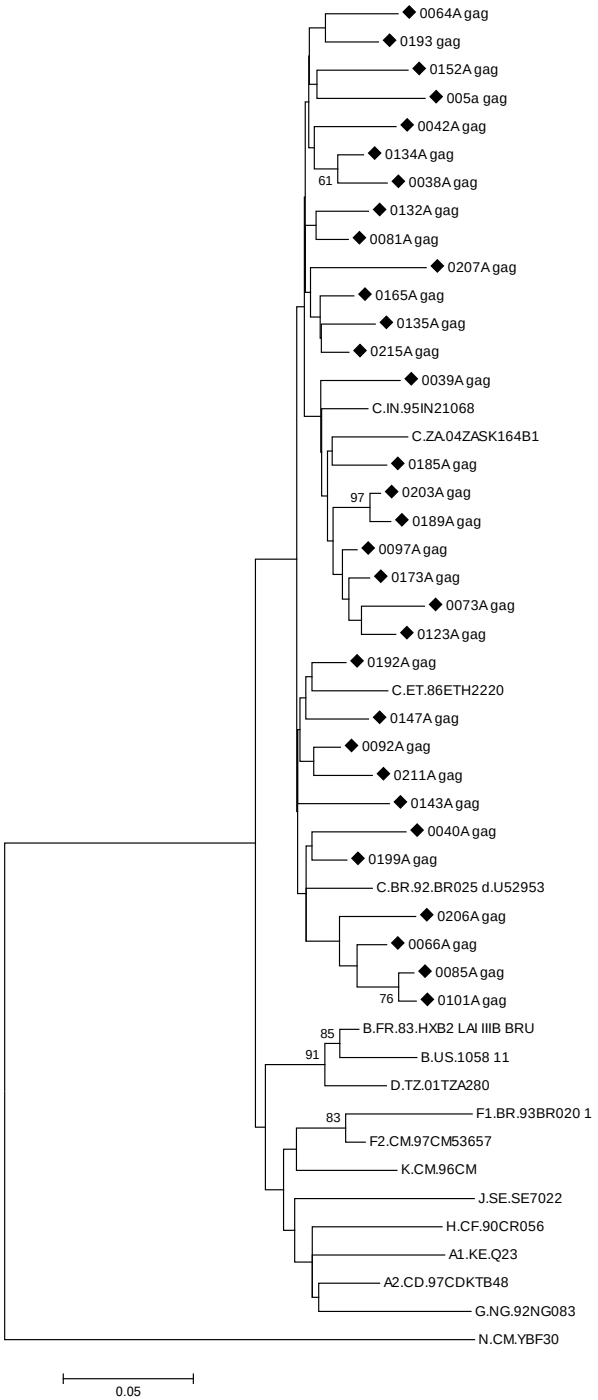



**Figure 3.12** Phylogenetic tree of the HIV-1 *gag* p24 genomic region. The Maximum Likelihood tree was constructed using the General Time Reversible evolutionary model implemented in MEGA version 5, plus discrete Gamma distribution was used to model evolutionary rate differences among sites. The bootstrap values of 1000 replicates above 70% are indicated next to the nodes and the scale bar at the bottom represents the number of base substitutions per site. The diamond indicates study samples.

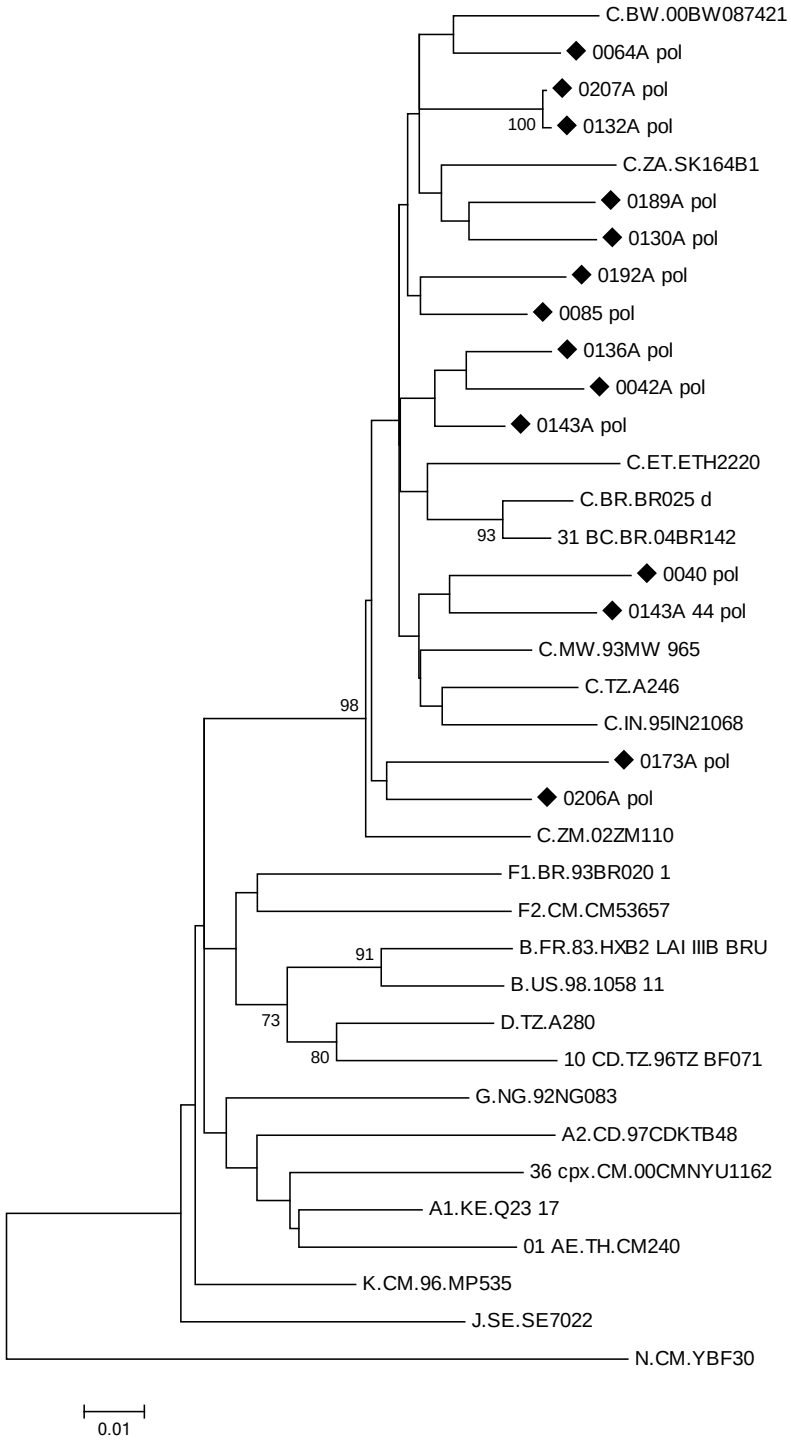



**Figure 3.13.** Phylogenetic tree of the HIV-1 *pol* p10 and p66/p51 genomic region. The Neighbor-Joining tree was constructed using the Kimura 2-parameter evolutionary model implemented in MEGA version 5. The bootstrap values of 1000 replicates above 70% are indicated next to the nodes and the scale bar at the bottom represents the number of base substitutions per site. Sequences 0207A and 0132A cluster together. The diamond indicates study samples.

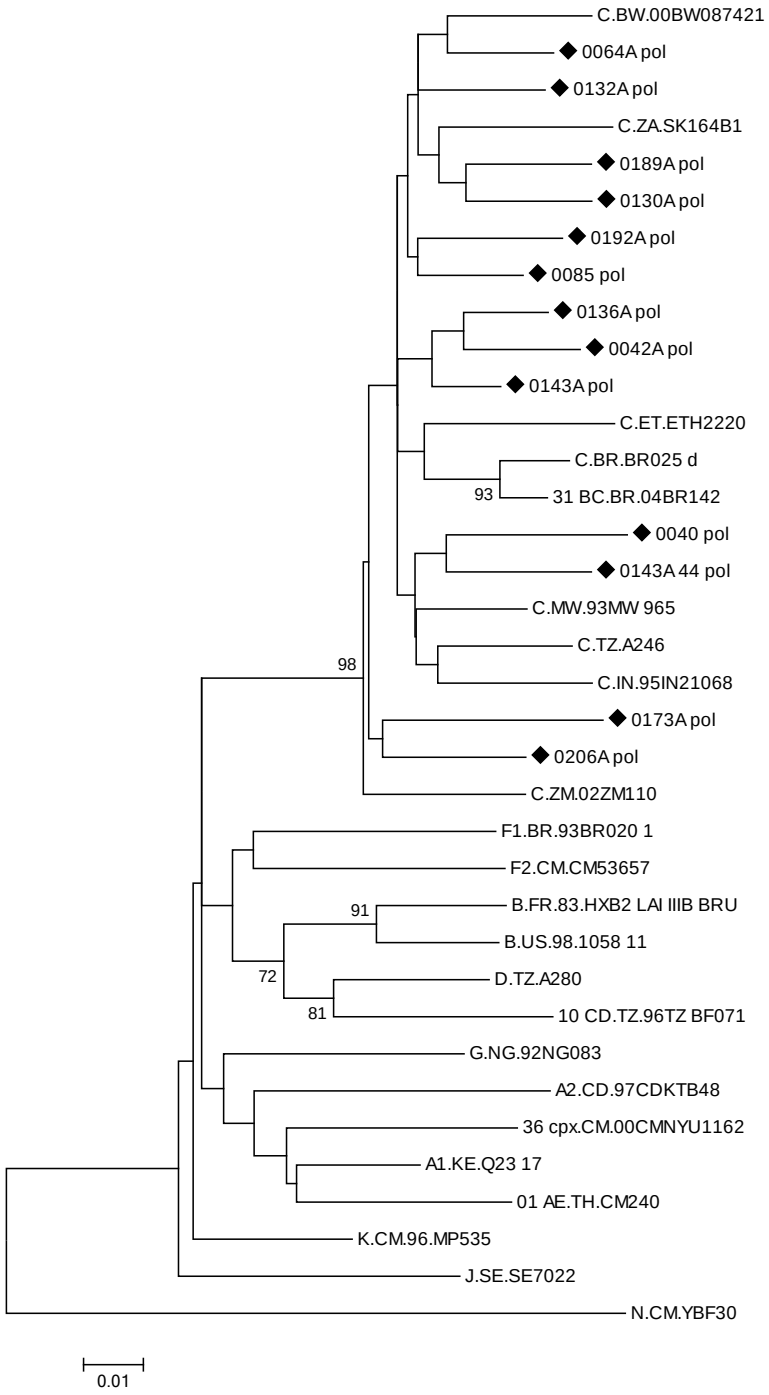



**Figure 3.14.** Phylogenetic tree of the HIV-1 *pol* p10 and p66/p51 genomic region. The Neighbor-Joining tree was constructed using the Kimura 2-parameter evolutionary model implemented in MEGA version 5. The bootstrap values of 1000 replicates above 70% are indicated next to the nodes and the scale bar at the bottom represents the number of base substitutions per site. The diamond indicates study samples.

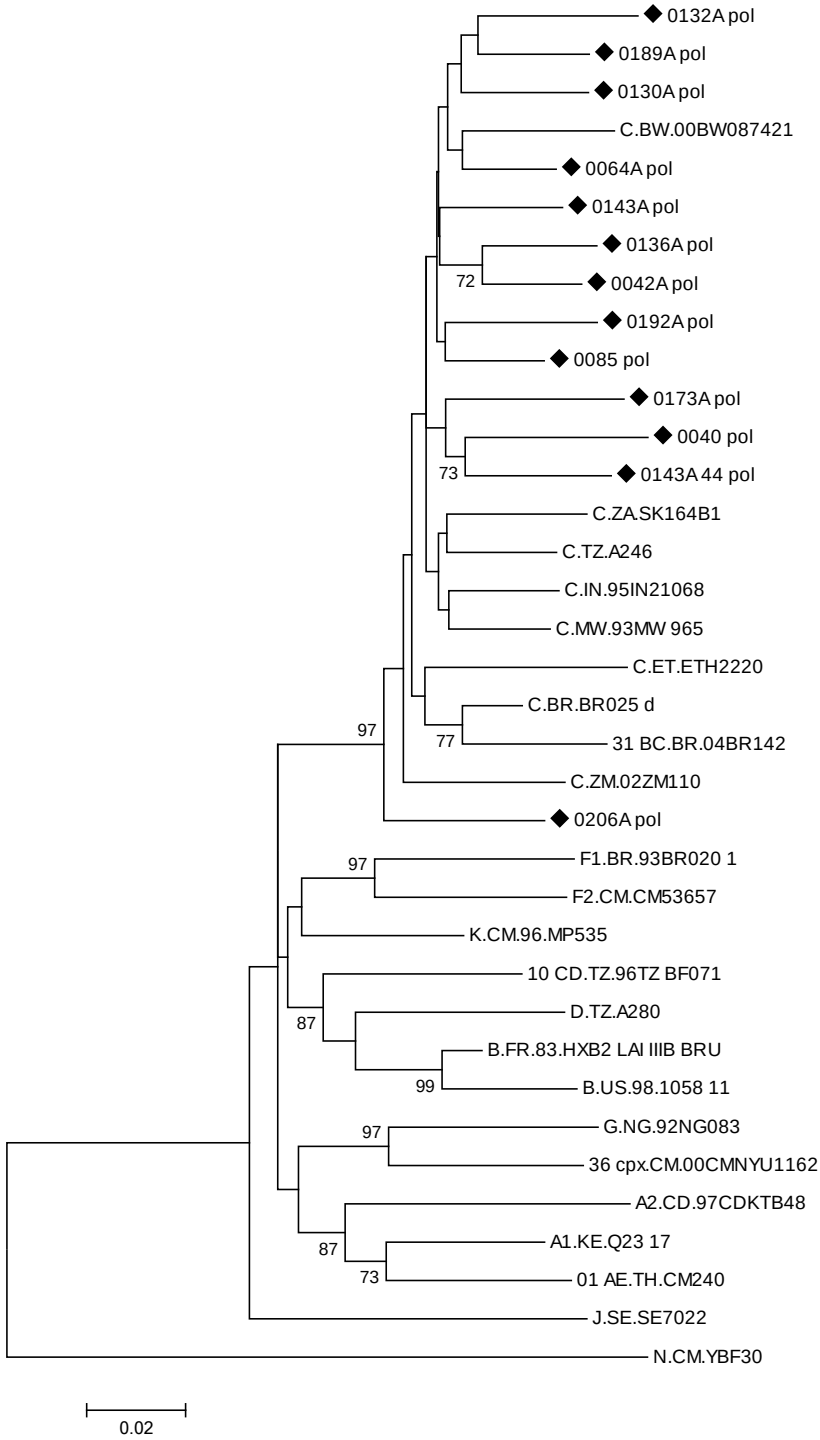



**Figure 3.15.** Phylogenetic tree of the HIV-1 *pol* p10 and p66/p51 genomic region. The Maximum Likelihood tree was constructed using the General Time Reversible evolutionary model implemented in MEGA version 5, plus discrete Gamma distribution was used to model evolutionary rate differences among sites. The bootstrap values of 1000 replicates above 70% are indicated next to the nodes and the scale bar at the bottom represents the number of base substitutions per site. The diamond indicates study samples.

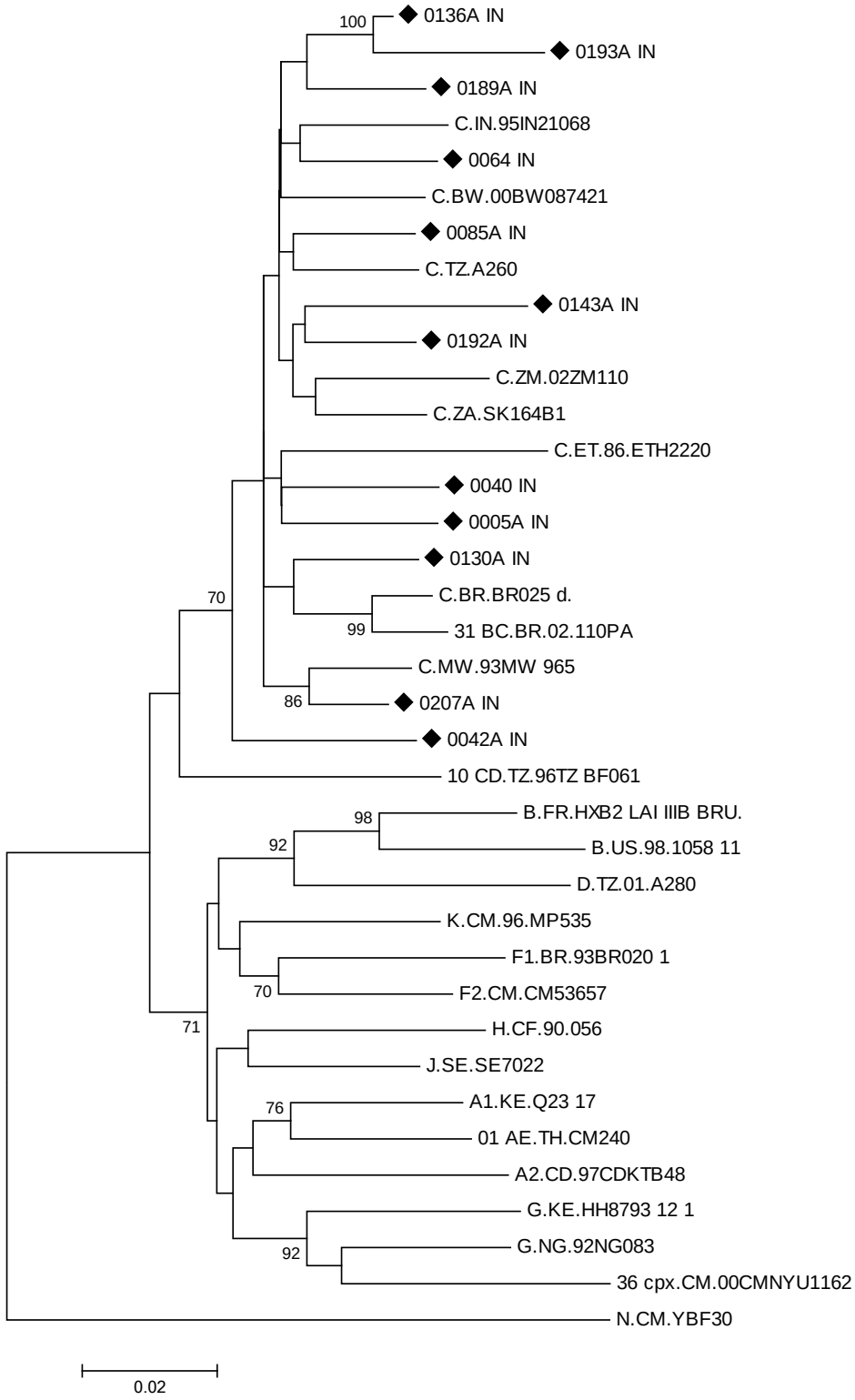



**Figure 3.16.** Phylogenetic tree of the HIV-1 *pol* p32 genomic region. The Neighbor-Joining tree was constructed using the Kimura 2-parameter evolutionary model implemented in MEGA version 5. The bootstrap values of 1000 replicates above 70% are indicated next to the nodes and the scale bar at the bottom represents the number of base substitutions per site. The diamond indicates study samples.

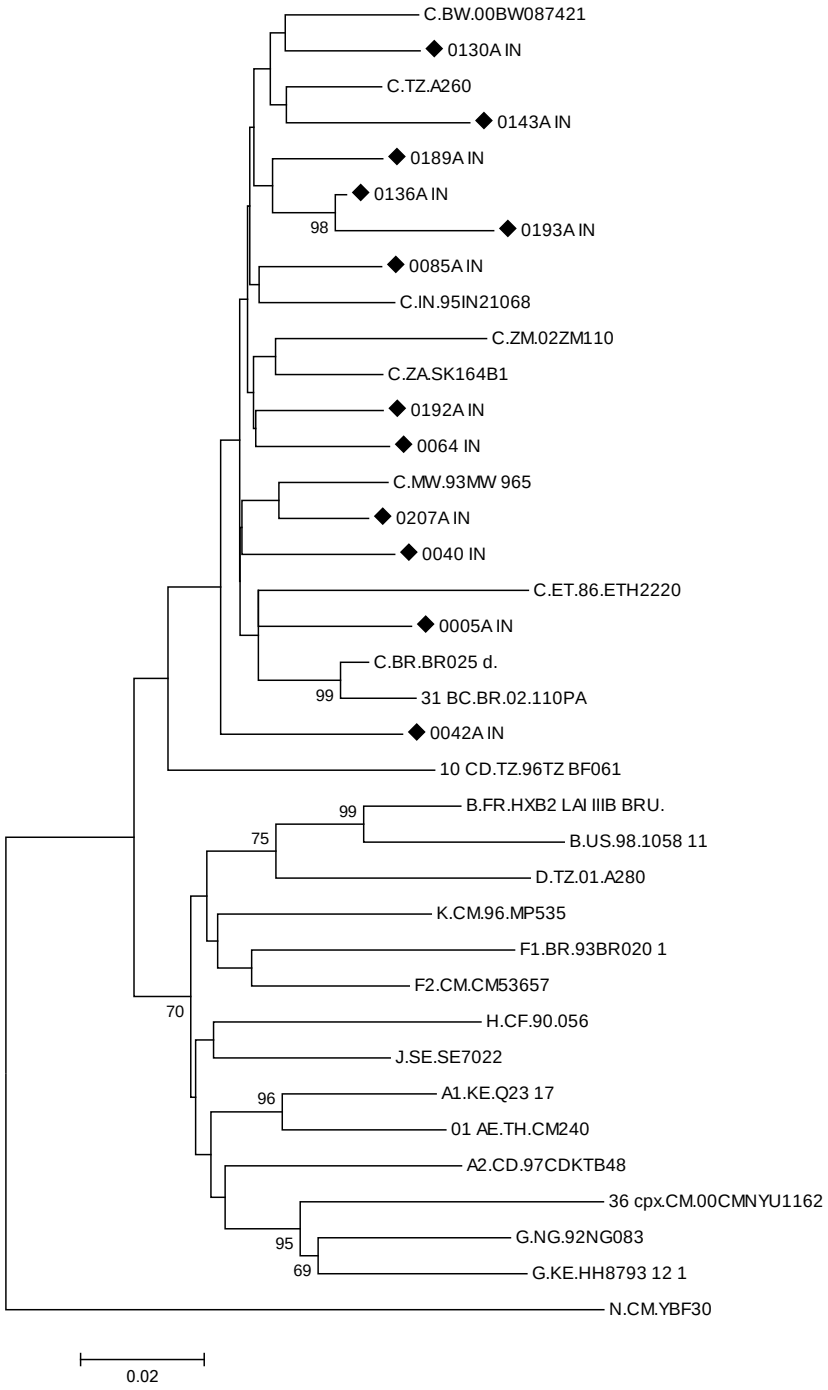



**Figure 3.17** Phylogenetic tree of the HIV-1 1 *pol* p32 genomic region. The Maximum Likelihood tree was constructed using the General Time Reversible evolutionary model implemented in MEGA version 5, plus discrete Gamma distribution was used to model evolutionary rate differences among sites. The bootstrap values of 1000 replicates above 70% are indicated next to the nodes and the scale bar at the bottom represents the number of base substitutions per site. The diamond indicates study samples.

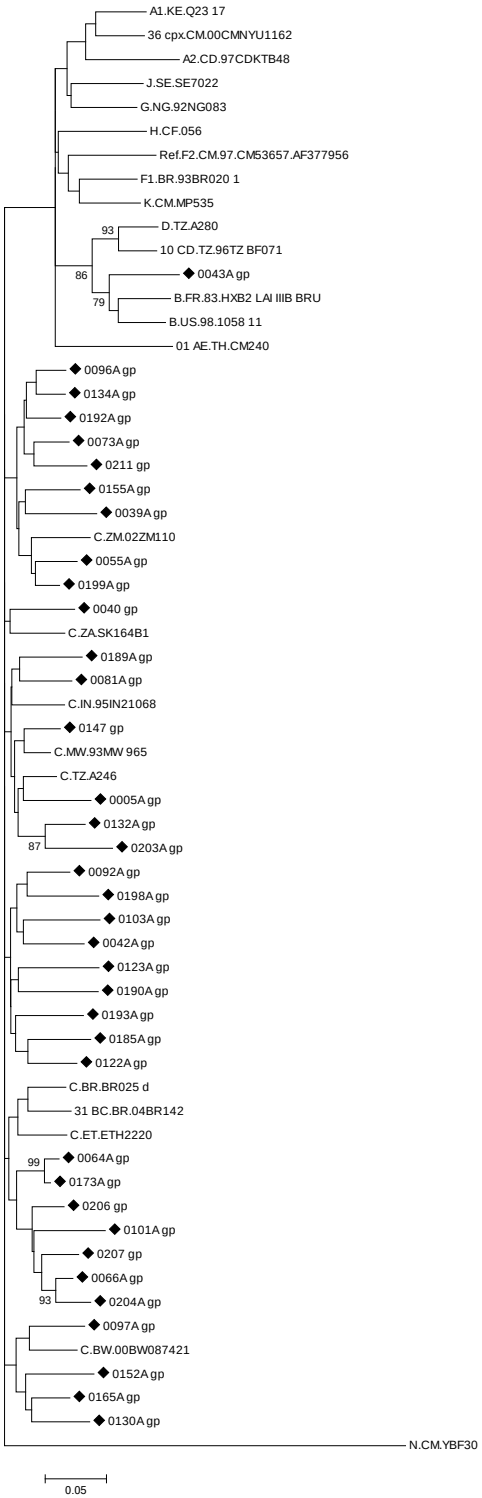



**Figure 3.18**. Phylogenetic tree of the HIV-1 *env* gp41genomic region. The Neighbor-Joining tree was constructed using the Kimura 2-parameter evolutionary model implemented in MEGA version 5. The bootstrap values of 1000 replicates above 70% are indicated next to the nodes and the scale bar at the bottom represents the number of base substitutions per site. The diamond indicates study samples.

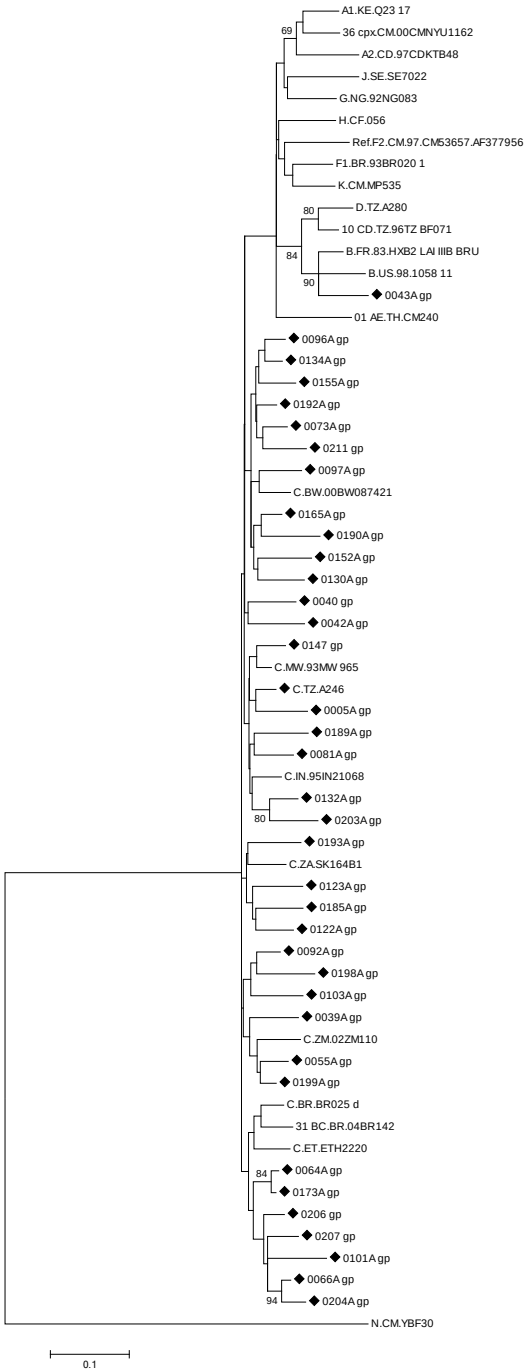



**Figure 3.19** Phylogenetic tree of the HIV-1 *env* gp41genomic region. The Maximum Likelihood tree was constructed using the General Time Reversible evolutionary model implemented in MEGA version 5, plus discrete Gamma distribution was used to model evolutionary rate differences among sites. The bootstrap values of 1000 replicates above 70% are indicated next to the nodes and the scale bar at the bottom represents the number of base substitutions per site. The diamond indicates study samples.

### 3.8 Transmitted drug resistance testing of the partial RT/PR and IN gene using CPR

Surveillance drug resistance mutations (SMDRs) detected in the 11 RT gene fragment derived from HIV-1 isolates from Bushbuckridge, Mpumalanga patient samples included the NRTI mutation T215S on the 0130A isolate, and the NNRTI mutation K103N on the 0143A isolate (**Table 3.5** and **Appendix**). However, no SMDRs were detected on the 13 PR and 11 IN gene fragments using the calibrated population resistance CPR tool. PCR amplification was successful for 29% of both the partial p66/p51 and p10 genes and 54.9% of the IN p32 gene. While the samples may not conform to some of the World Health Organisation's eligibility criteria for HIVDR surveys, including the requirement for the $CD4^+$ T cells count of > 500 cells/ml, they do satisfy some of the criteria such as sample size and a small geographic area.

#### 3.8.1 Drug resistance mutations and polymorphisms detected using the HIVdb.

**Table 3.5.** HIV-1 drug resistance mutations and polymorphisms against NNRTIs and PIs, respectively.

| Isolate | Major PI Mutations | Minor PI Mutations | Other PI mutations | NRTI mutations | NNRTI mutations |
|---|---|---|---|---|---|
| 0040A | — | — | M36I | — | — |
| 0042A | — | — | M36I | — | — |
| 0064A | — | — | M36I | — | — |
| 0085A | — | — | M36I | — | — |
| 0130A | — | — | M36I | T215S | — |
| 0132A | — | — | M36I | — | K101Q |
| 0136A | — | — | M36I | — | — |
| 0143A | — | — | M36I | — | — |
| 0143A_44 | — | — | M36I | — | K103N E138A |
| 0173A | — | — | M36I T74S | — | — |
| 0189A | — | — | M36I | — | K101E |
| 0192A | — | — | M36I T74S | — | — |
| 0204A | — | L90R | M36I | — | — |
| 0206A | — | — | T74S | — | — |
| 0207A | — | — | M36I | — | K101Q |

The T215S mutation does not decrease NRTI susceptibility but arises from viruses that once contained T215Y/F. The T215Y/F cause AZT and D4T

resistance and reduce susceptibility to ABC, ddI, and TDF. The NNRTI mutation K103N detected on the 0143A_44 isolate causes high-level resistance to NVP, and EFV. The NNRTI mutation K101Q, which was also detected on the the 0132A and 0207A isolates, is a relatively nonpolymorphic mutation that occurs slightly more commonly in patients receiving NNRTIs, and may contribute to reduced NVP, EFV, and ETR susceptibility when present with other NNRTI-resistance mutations. The NNRTI mutation, E138A, detected on the 0143A_44 isolate is a polymorphism that may contribute to reduced ETR and RPV susceptibility in combination with other NNRTI-resistance mutations. The K101E mutation found on the 0189A isolate causes intermediate resistance to NVP and low-level resistance to EFV, ETR, and RPV. Except for the L90R minor PI mutation detected on the 0204A isolate, no major or minor PI mutations were detected in the Bushbuckridge, Mpumalanga isolates. The L90M mutation reduces susceptibility to NFV, SQV/r, ATV/r, and IDV/r when present with other mutations. One of the most common mutations, M36I, a consensus amino acid in most non-B subtypes, is weakly associated with PI resistance in subtype B viruses when present with other mutations. The T74S mutation occurs in about 5% of untreated persons with subtype C viruses and is associated with reduced NFV susceptibility (Barth et al., 2008; Cane et al., 2001; Soares et al., 2009; Velasquez-Campoy et al., 2001).

# CHAPTER FOUR

# DISCUSSION AND CONCLUSION

# CHAPTER FOUR

# DISCUSSION AND CONCLUSIONS

## 4. 1. Discussion

The investigation of the HIV subtype diversity of samples obtained from a cohort in Bushbuckridge, Mpumalanga revealed, first, that the HIV from these samples belong almost entirely to HIV-1 subtype C; second, that the way in which the sequences derived from these samples cluster in phylogenetic trees suggests there has been multiple introductions of HIV into Bushbuckridge; and third, the prevalence of SDRMs, as well as antiretroviral drug resistance mutations and drug resistance-associated polymorphisms, in Bushbuckridge is extremely low.

### 4.1.1 Bushbuckridge HIV epidemic is predominantly subtype C.

Phylogenetic analysis revealed that the partial sub-genomic *gag* p24, *pol* p10 and p66/p51, *pol* p32 and *env* gp41 sequences derived from patients from Bushbuckridge, Mpumalanga, belong to HIV-1 subtype C. The Bushbuckridge, Mpumalanga sequences almost invariably clustered with HIV-1 subtype C reference sequences from South Africa. Botswana, Malawi, Zambia, India, Ethiopia and Brazil in both NJ and ML trees constructed using MEGA. The only sequence that clustered with HIV-1 subtype B sequences was *env* gp41 0043A_gp. However, the PCR amplification of the *gag* p24, *pol* p10 and p66/p51, *pol* p32 genomic regions of the 0043A sample were not successful, which implies they may not necessarily have clustered with HIV-1 subtype B sequences, and that the 0043A sample could have been a recombinant.

The HIV-1 subtyping of the *gag* p24, *pol* p10 and p66/p51, *pol* p32 and *env* gp41 sequences using the LANL QC RIP tool, jpHMM and REGA, as well as

subtyping using SCUEAL for *pol* sequences, revealed that most of the Bushbuckridge, Mpumalanga isolates belonged to subtype C. Subtype analysis of the Bushbuckridge sequences using jpHMM and REGA subtyping tools also detected recombination between sequence segments from different HIV-1 subtypes within single subgenomic fragments in the 0042A_gag, 0132A_pol, 0189A_gag, 0192A_IN, 0193A_IN and 0203_gag sequences.

The fact that the HIV samples from Bushbuckridge, Mpumalanga, belong almost entirely to subtype C is consistent not only with the explosive HIV-1 epidemic in southern Africa, but also it's very limited subtype diversity. HIV-1 subtype C is the most common subtype, accounting for 92% of HIV infections in southern Africa, while subtype B is responsible for about 7% of infections. (Butler et al., 2007; Bredell., et al., 1998; Bredell et al., 2000; Bredell et al., 2002; Morris et al., 2000; Novitsky et al., 1999; Van Harmelen et al.; Williamson et al.; UNAIDS, 2012).

### 4.1.2 Multiple introductions of HIV into Bushbuckridge.

Significantly, the Bushbuckridge isolates do not only cluster with the SouthAfrican HIV-1 subtype C isolate in both the NJ and ML phylogenetic trees, but also cluster and intermingle with the radiation comprising HIV-1 subtype C isolates from Botswana, Malawi, Zambia, as well as India. That is the sequences from South Africa and southern African countries, as well as the one from India (IN), do not exhibit country-specific clusters and instead show a spread of the sequences throughout the tree indicating multiple introductions of subtype C viruses in these countries. The fact that HIV-1 subtype C sequences from South Africa tend to intermingle with HIV-1 subtype C sequences from Botswana, Malawi and Zambia suggests they may have a common evolutionary origin. The possibility of an underlying common evolutionary origin of isolates in southern Africa is consistent with the history of the population dynamics of the southern African region. While the HIV-1 subtype C isolates from Brazil and Ethiopia tend to cluster separately, the fact that the subtype C isolate from India tends to cluster with the subtype C isolates from southern Africa can be explained by the historical connections between the Indian subcontinent and southern Africa arising from the roles of

both regions as former British colonial territories. Thus, there may be continuing movement of the virus between Africa and India. As a consequence, a strain like IN101 (accession number AB023804) from India is closer to most African subtype C strains than African strains are to each other (Gaschen et al., 2002; Shankarappa et al., 2001).

One of the central elements of the population dynamics of not only South Africa but the broader southern African region, critical in shaping the patterns of population mobility and integration characterizing the region, was indubitably the migrant labour sytem. The migrant labour system, which linked countries in southern Africa in which adult national HIV prevalence exceeded 15% in 2007, was integral to the development and structure of the South African economy and apartheid. Botswana, Lesotho, Namibia, South Africa, Swaziland, Zambia, and Zimbabwe, were historically linked through the migrant labour system that brought men from as far as Zambia and Malawi to the mines initially on the Reef and subsequently elsewhere in the country (Abdool Karim and Abdool Karim, 2002; Bauer and Taylor, 2005; Dusheiko et al., 1989; Huang et al., 2009).). Migrants are more vulnerable to HIV infection than are people who do not move, both in southern Africa as in other African countries (Abdool Karim, 1992; Decosas, 1995; Lurie, 2001). A 1985 survey of workers in the gold mines from the whole southern African region showed HIV to be rare in South African miners but already at 3% prevalence in Malawians (Abdool Karim and Abdool Karim, 2002). High infection levels are being found in Gaza province in Mozambique, where large numbers of migrants working in South Africa originate (Ministry of Health, Mozambique, 2005). Before and after independence foreign migrant workers also crossed borders to work in mines in Namibia, Botswana, Zambia, and Zimbabwe (Bauer and Taylor, 2005; Lurie, 2000; Lurie et al., 2003).

The fact that many of the southern African countries with the explosive HIV/AIDS epidemic are landlocked entails that the region’s road transport networks does not just link these landlocked countries to the ports in Durban, Richards Bay and Maputo, but also facilitate the rapid spread of HIV in the region by ensuring the sexual networks that drive the epidemic transcend

national boundaries. The Ehlanzeni District in Mpumalanga Province straddles the Maputo Corridor, a major trade route which connects the Gauteng, Limpopo, and Mpumalanga provinces of South Africa with Maputo, the capital of Mozambique that also has a major port. In Mozambique, HIV is spreading more rapidly in provinces linked by major transport routes to Malawi, South Africa and Zimbabwe. High infection rates have been found in Sofala province, which is traversed by Zimbabwe's main export route (Mozambique, 2005; Ramjee and Gouws, 2002).

The peculiarly explosive HIV-1 epidemic in southern Africa could also stem from the unique biological properties of subtype C.HIV-1 subtype C has an additional NF-κB binding site in the long terminal repeat (LTR), a prematurely truncated Rev protein, a 5′-amino-acid insertion in Vpu, and a more active, catalytically efficient protease, which may influence viral gene expression and alter the transmissibility and pathogenesis of subtype C isolates (De Oliveira et al., 2003; Gao et al., 1998; Hunt and Tiemessen, 2000; McCormick-Davis et al., 2000; Rodenburg et al., 2001; Valesquez-Campoy et al., 2001). These unique biological properties, including those related to viral entry and pathogenesis such as the CCR5 and non-syncytium-inducing phenotype, may account for the explosive epidemic of HIV-1 subtype C in southern Africa (Ball et al., 2003; Peeters et al., 1999; Ping et al., 1999). However, the additional NF-κB site in HIV-1 subtype C may be biologically inactive, and enhanced activity of these individual functions may still not be sufficient to overcome the decreased replicative capacity of the CCR5-tropic non-syncytium-inducing phenotype (Ball et al., 2003).

### 4.1.3 SMDRs, resistance mutations and polymorphisms in the cohort.

Two SMDRs, T215S and K103N detected on the 0130A and 0143_44 isolates respectively, were also found on the same isolates using the Stanford University Drug Resistance Database. The T215S and K103N mutations suggest the patients may either have undergone antiretroviral treatment or been infected with antiretroviral drug resistant strains (Barth et al., 2008; Cane et al., 2001; Soares et al., 2009; Velasquez-Campoy et al., 2001).

### 4.1.4 Limitations of the study.

The limitations of the study include a relatively small sample size; DNA amplification was not successful for up to 71% of the samples of, for example, the partial pol p10 and p66/p51 sub-genomic region; use of partial gene regions to assign viral subtypes, potentially allowing recombinant viruses to be missed, the use of direct; population sequencing may result in the lack of detection of minority-population viruses; which can lead to an underestimation of viral diversity and drug resistance mutations (Iweriebor et al., 2012).

## 4.2. Conclusions

HIV diversity may have implications for diagnosis, pathogenesis, transmission, clinical management and vaccine development. Phylogenetic analysis of HIV sequence diversity has allowed vital insights into the origin, evolution and spread of HIV, which suggests it is imperative to maintain HIV molecular epidemiology surveillance. Molecular epidemiological investigations in South Africa have, with the notable exceptions of the Free State and Limpopo, largely focused on provinces with major metropolitan centers such as Guateng, the Western Cape and KwaZulu Natal. Hardly any molecular epidemiological studies have been undertaken in the Eastern Cape, Mpumalanga, North West and Northern Cape provinces. The extensive population mobility arising from the historical and structural migrant labour system characterizing South Africa, and the concomitant overlapping of sexual networks, seems to have precluded the possibility of distinct geographical lineages developing. However, the demise of apartheid, in particular the end of influx control measures, may have a significant impact on patterns of population mobility and settlement in South Africa, which in turn may affect the patterns of transmission of HIV and ultimately it's evolution. The possible emergence of various HIV-1 recombinants could suggest that the migration of people into South Africa from Central, West and eastern Africa could also impact on the character and dynamics of the HIV/AIDS epidemic in South Africa (Engelbrecht et al., 1999).

# CHAPTER FIVE

# REFERENCES

Abreu, C. M., Brindeiro, P. A., Martins, A. N., Arruda, M. B., Bule, E., Stakteas, S., Tanuri, A., Brindeiro, R. de M., 2008. Genotypic and phenotypic characterization of human immunodeficiency virus type 1 isolates circulating in pregnant women from Mozambique. Arch. Virol.; 153: 2013-2017.

Abdool Karim, Q., Abdool Karim, S. S., 2002. The evolving HIV epidemic in South Africa. Int. J. Epidemiol.; 31: 37-40.

Alaeus, A., Lidman, K., Björkman, A., Giesecke, J., Albert, J., 1999. Similar rate of disease progression among individuals infected with HIV-1 genetic subtypes A-D. AIDS; 13: 901-907.

Alaeus, A., Lidman, K., Sönnerborg, A., Albert, J., 1997. Subtype-specific problems with quantification of plasma HIV-1 RNA. AIDS; 11: 859-865.

Alfaro, M. E., Holder, M. T., 2006. The posterior and the prior in Bayesian phylogenetics. Annu. Rev. Ecol. Evol. Syst.; 37: 19-42.

Alfaro, M. E., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol.; 20: 255-266.

Anderson, J. P., Learn, G. H., Rodrigo, A. G., He, X., Wang, Y., Weinstock, H., Kalish, M. L., Robbins, K. E., Hood, L., Mullins, J. I., 2003. Predicting demographic group structures based on DNA sequence data. Mol. Biol. Evol., 20: 1168-1180.

Apetrei, C., Marx, P. A., Smith, S. M., 2004. The evolution of HIV and its consequences. Infect. Dis. Clin. N. Am.; 18: 369-394.

Ariën, K. K., Vanham, G., Arts, E. J., 2007. Is HIV-1 evolving to a less virulent form in humans? Nat. Rev. Micriobiol.; 5: 141-151.

Ariyoshi, K., Matsuda, M., Miura, H., Tateishi, S., Yamada, K., Sugiura, W., 2003. Patterns of point mutations associated with antiretroviral drug treatment failure in CRF01_AE (subtype E) infection differ from subtype B infection. J. Acquir. Immune Defic. Syndr.; 33: 336-342.

Arora, S. K., Gupta, S., Toor, J. S., Singla, A., 2008. Drug resistance-associated genotypic alterations in the *pol* gene of HIV type 1 isolates in ART-naive individuals in North India. AIDS Res. Hum. Retrovir.; 24: 125-130.

Arroyo, M. A., Hoelscher, M., Sanders-Buell, E., Herbinger, K. H., Samky, E., Maboko, L., Hoffmann, O., Robb, M. R., Birx, D. L. McCutchan, F. E., 2004. HIV type 1 subtypes among blood donors in the Mbeya region of southwest Tanzania. AIDS Res. Hum. Retrovir.; 20: 895-901.

Ataher, Q., Portsmouth, S., Napolitano, L. A., Eng, S., Greenacre, A., Kambugu, A., Wood, R., Badal-Faesen, S., Tressler, R., 2012. The epidemiology and clinical correlates of HIV-1 co-receptor tropism in non-subtype B infections from India, Uganda and South Africa. J. International AIDS Society; 15: 2 doi: 10.1186/1758-2652-15-2.

Baeten, J. M., Chohan, B., Lavreys, L., Chohan, V., McClelland, R. S., Certain, L., Mandaliya, K., Jaoko, W., Overbaugh, J., 2007. HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. J. Infect. Dis.; 195: 1177-1180.

Baldauf, S. L., 2003. Phylogeny for the faint of heart: a tutorial. Trends in Genetics; 19: 345-351.

Ball, S. C., Abraha, A., Collins, K. R., Marozsan, A. J., Baird, H., Quiñones-Mateu, M. E., Penn-Nicholson, A., Murray, M., Richard, N., Lobritz, M., Zimmerman, P. A., Kawamura, T., Blauvelt, A., Arts, E. J., 2003. Comparing the *ex vivo* fitness of CCR5- tropic human immunodeficiency virus type 1 isolates of subtypes B and C. J. Virol.; 77: 1021-1038.

Barré-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Chamalet, S., Gruest, J., Dauguet, C., Axler-Blin, C., 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science; 220: 868-871.

Barth, R. E., Wensing, A. M., Tempelman, H. A., Moraba, R., Rob Schuurman, R., Hoepelman, A. I., 2008. Rapid accumulation of nonnucleoside reverse transcriptase inhibitor-associated resistance: evidence of transmitted resistance in rural South Africa. AIDS, 22: 2210-2212.

Bauer, G., Taylor, S. D., 2005. Politics in Southern Africa: state and society in transition. Lynne Rienner Publishers, Boulder, Colorado, USA.

Becker, M. L. B., De Jager, G., Becker, W. B., 1995. Analysis of partial *gag* and *env* gene sequences of HIV type 1 strains from southern Africa. AIDS Res. Hum. Retrovir.: 11: 1265-1267.

Beer, B. E., Bailes, E., Sharp, P. M., Hirsch, V. M., Diversity and evolution of primate lentiviruses. In: Kuiken, C. L., Foley, B., Hahn, B., *et al*. (eds) AIDS and Human Retroviruses HIV Sequence Database, part III. Los Alamos, New Mexico: Theoretical Biology and Biophysics, Los Alamos National Laboratory, 1999 (http://hiv-web.lanl.gov)

Bennett, D. E., Camacho, R. J., Otelea, D., Kuritzkes, D. R., Fleury, H., Kiuchi, M., Heneine, W., Kantor, R., Jordan, M. R., Schapiro, J. M., Vandamme, A-M., Sandstrom, P., Boucher, C. A. B., Van de Vijver, D., Rhee, S-Y., Liu, T. F., Pillay, D., Shafer, R. W., 2009. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. PLoS ONE; 4: e4724.

Bernardin, F., Kong, D., Peddada, L., Baxter-Lowe, L. A. Delwart, E., 2005. Human immunodeficiency virus mutations during the first month of infection are preferentially found in known cytotoxic T-lymphocyte epitopes. J. Virol.; 79: 11523-11528.

Bessong, P. O., Iweriebor, B. C., Mavhandu, L. G., Masebe, T., Nwobegahay, J. 2008. Tools for genotyping human immunodeficiency virus, and implications of genetic diversity on diagnostics, treatment and prevention. African J. Biotechnology: 7: 4774-4783.

Bessong, P. O., Mphahlele, J., Choge, I. A., Obi, L. C., Morris, L., Hammarskjold, M-L., Rekosh, D. M., 2006. Resistance mutational analysis of HIV type 1 subtype C among rural South African drug-naive patients prior to large-scale availability of antiretrovirals. AIDS Res. Human Retrovir.; 22: 1306-1312.

Bessong, P. O., Obi, C. L., Cilliers, T., Choge, I., Phoswa, M., Pillay, C., Papathanasopoulos, M., Morris, L., 2005. Characterization of human immunodeficiency virus type 1 from a previously unexplored region of South Africa with a high HIV prevalence. AIDS Res. Human Retrovir.; 21: 103-109.

Björndal, Å., Deng, H., Jansson, M., Fiore, J., Colognesi, C., Karlsson, A., 1997. Coreceptor usage of primary HIV-1 isolates varies according to biological phenotype. J. Virol.; 71: 7478-7487.

Bonhoeffer, S., Holmes, E. C., Nowak, M. A., 1995. Causes of HIV diversity. Nature; 376; 125.

Bredell, H., Hunt, G., Casteling, A., Cilliers, T., Rademeyer, C., Coetzer, M., Miller, S., Johnson, D., Tiemessen, C. T., Martin, D. J., Williamson, C., Morris, L., 2002. HIV-1 Subtype A, D, G, AG and unclassified sequences identified in South Africa. AIDS Res. Human Retrovir.; 18: 681-683.

Bredell, H., Hunt, G., Morgan, B., Tiemessen, C. T., Martin, D. J., Morris, L., 2000. Identification of HIV type 1 intersubtype recombinants in South Africa using *env* and *gag* heteroduplex mobility assays. AIDS Res. Hum. Retrovir.; 16: 493-497.

Bredell, H., Williamson, C., Sonnenberg, P., Martin, D. J., Morris, L., 1998. Genetic characterization of HIV type 1 from migrant workers in three South African gold mines. AIDS Res. Human Retrovir.; 14: 677-684.

Brenner, B. G., Oliveira, M., Doualla-Bell, F., Moisi, D. D., Ntemgwa, M., Frankel, F., Essex, M., Wainberg, M. A., 2006. HIV-1 subtype C viruses rapidly develop K65R resistance to tenofovir in cell culture. AIDS; 20: F9-F13.

Brenner, B., Turner, D., Oliveira, M., Moisi, D., Detorio, M., Carobene, M., Marlink, R. G., Schapiro, J., Roger, M., Wainberg, M. A., 2003. A V106M mutation in HIV-1 clade C viruses exposed to efavirenz confers cross-resistance to non-nucleoside reverse transcriptase inhibitors. AIDS; 17: F1-F5.

Briggs, J. A.G., Wilk, T., Welker, R., Kräusslich, H-G.,.Fuller, S. D., 2003. Structural organization of authentic, mature HIV-1 virions and cores. EMBO J.; 22: 1707-1715.

Brown, B. K., Darden, J. M., Tovanabutra, S., Oblander, T., Frost, J., Sanders-Buell, E., De Souza, M. S., Birx, D. L., McCutchan, F. E. Polonis, V. R. 2005. Biologic and genetic characterization of a panel of 60 human immunodeficiency virus type 1 isolates, representing clades A, B, C, D, CRF01_AE, and CRF02_AG, for the development and assessment of candidate vaccines. J. Virol.; 79: 6089-6101.

Butler, I.F., Pandrea, I., Marx, P.A., Apetrei, C., 2007. HIV genetic diversity: biological and public health consequences. Curr. HIV Res. 5, 23-45.

Cane, P. A., De Ruiter, A., Rice, P., Wiselka, M., Fox, R., Pillay, D., 2001. Resistance associated mutations in the human immunodeficiency virus type 1 subtype C protease gene from treated and untreated patients in the United Kingdom. J. Clin. Microbiol.; 39: 2652-2654.

Cann, A. J., Churcher, M. J., Boyd, M., O'Brien, W., Zhao, J. Q., et al, 1992. The region of the envelope gene of human immunodeficiency virus type 1 responsible for determination of cell tropism. J. Virol.; 66: 305-309.

Cao, H., Kanki, P., Sankale, J. L., Korber, B., M'Boup, S., Walker, B. D., 1997. CTL cross-reactivity among different HIV-1 clades: implications for vaccine development. J. Virol.; 71: 8615-8623.

Carr, J. K., Torimiro, J. N., Wolfe, N. D., Eitel, M. N., Kim, B., Sanders-Buell, E., Jagodzinski, L. L., Gotte, D., Burke, D. S., Birx, D. L., McCutchan, F. E., 2001. The AG recombinant IbNG and novel strains of group M HIV-1 are common in Cameroon. Virology; 286: 168-181.

Carr, J. K., Foley, B. T., Leitner, T., Salminen, M., Korber, B., McCutchan, F., 1998. Reference sequences representing the principal genetic diversity of HIV-1 in the pandemic.

Carr, J. K., Salminen, M. O., Koch, C., Gotte, D., Artenstein, A. W., Hegerich, P. A., St Louis, D., Burke, D. S., McCutchan, F. E. 1996. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. J. Virol.; 70: 5935-5943.

CDC, 1987. Human immunodeficiency virus infection in the United States. Morbidity Mortality Weekly Rep.; 36B: 801-804. (http://www.cdc.gov/mmwr/mmwr_wk/wk_cvol.html) (Accessed: 6 March 2012).

CDC, 1982a. Update on Kaposi's sarcoma and opportunistic infections in previously healthy persons- United States. Morbidity Mortality Weekly Rep.; 31: 294-301. (http://www.cdc.gov/mmwr/mmwr_wk/wk_cvol.html) (Accessed: 6 March 2012).

CDC, 1982b. Epidemiologic notes and reports persistent, generalized lymphadenopathy among homosexual males. Morbidity Mortality Weekly Rep.; 31: 249-251. (http://www.cdc.gov/mmwr/mmwr_wk/wk_cvol.html) (Accessed: 6 March 2012).

CDC, 1981. Kaposi's sarcoma and *Pneumocystis pneumonia* among homosexual men - New York City and California. Morbidity Mortality Weekly Rep; 30: 305-308. (http://www.cdc.gov/mmwr/mmwr_wk/wk_cvol.html) (Accessed: 6 March 2012).

Chen, J., Dang, Q., Unutmaz, D., Pathak, V. K., Maldarelli, F., Powell, D., Hu, W-S., 2005. Mechanisms of nonrandom human immunodeficiency virus type 1

infection and double infection: preference in virus entry is important but is not the sole factor. J. Virol.; 79: 4140-4149.

Coffin, J. 1999. Molecular biology of HIV. 3-40. Editor: Crandall, K.A., 1999. The Evolution of HIV. Johns Hopkins University Press, Baltimore, MD.

Coovadia, H., Jewkes, R., Barron, P., Sanders, D., McIntyre, D., 2009. The health and health system of South Africa: historical roots of current public health challenges. Lancet; 374: 817-834.

Craigie, R., Bushman, F. D., 2012. HIV DNA integration. Cold Spring Harb. Perspect. Med.; 2: a006890.

Decosas, J., Kane, F., Anarfi, J. K., Sodji, K. D. R., Wagner, H. U. 1995. requirements Migration and AIDS. Lancet; 346: 826-828.

De Jong, J. J., De Ronde, A., Keulen, W., Tersmette, M., Goudsmit, J., 1992. Minimal for the human immunodeficiency virus type 1 domain to support the syncytium inducing phenotype: analysis by single amino acid substitution. J. Virol.; 66: 6777-6780.

Delport, W., Poon, A. F. Y., Frost, S. D. W., Kosakovsky Pond, S. L., 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics; 26: 2455-2457.

De Oliveira, T., Deforche, K., Cassol, S., Salminen, M., Paraskevis, D., Seebregts, C., Snoeck, J., Janse van Rensburg, E., Annemarie M. J. Wensing, A. M. J., Van de Vijver, D. A., Boucher, C. A., Camacho, R., Vandamme, A-M., 2005. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics; 21: 3797-3800.

De Oliveira, T., Engelbrecht, S., Van Rensburg, E., Gordon, M., Bishop, K., Zur Megede, J., Barnett, S. W., Cassol, S., 2003. Variability at HIV- 1 subtype C protease cleavage sites and indication of viral fitness? J. Virol.; 77: 9422-9430.

Djoko, C. F., Rimoin, A. W., Vidal, N., Tamoufe, U., Wolfe, N. D., Butel, C., LeBreton, M., Tshala, F. M., Kayembe, P. K., Muyembe, J. J., Edidi-Basepeo, S., Pike, B. L., Fair, J. N., Mbacham, W. F., Saylors, K. E., Mpoudi-Ngole, E., Delaporte, E., Grillo, M., Peeters, M., 2011. High HIV type 1 group M *pol* diversity and low rate of antiretroviral resistance mutations among the uniformed services in Kinshasa, Democratic Republic of the Congo. AIDS Res. Hum. Retrovir.; 27: 323-329.

Dusheiko, G. M., Brink, B. A., Conradie, J. D., Marimuthu, T., Sher, R., 1989. Regional prevalence of hepattis B, delta, and human immunodeficiency virus infection in southern Africa: a large population survey. Am. J. Epidemiol.; 129: 138-145.

Efron, B., Halloran, E., Holmes, S., 1996. Bootstrap confidence levels for phylogenetic trees. Proc. Natl. Acad. Sci. USA. 93:13429–13434.

Egan, A. N., Crandall, K. A., Theory of phylogenetic estimation. 1-26. (Unpublished).

Ellenberger, D. L., Sullivan, P. S., Dorn, J., Schable, C., Spira, T. J., Folks, T. M., Lal, R. B., 1999. Viral and immunologic examination of human immunodeficiency virus type 1-infected, persistently seronegative persons. J. Infect. Dis.; 180: 1033-1042.

Engelbrecht, S., Smith, T.-L., Kasper, P., Faatz, E., Zeier, M., Moodley, D., Clay, C. G., Van Rensburg, E. J., 1999. HIV type 1 V3 domain serotyping and genotyping in Gauteng, Mpumalanga, KwaZulu-Natal, and Western Cape Provinces of South Africa. AIDS Res. Human Retrovir.; 15: 325-328.

Engelbrecht, S., Laten, J. D., Smith, T-L., Van Rensburg, E. J., 1995. Identification of *env* subtypes in fourteen HIV type 1isolates from South Africa. AIDS Res. Human Retrovir.; 11: 1269-1271.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution; 39: 783-791.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a Maximum Likelihood approach. J. Mol. Evol.; 17: 368-376.

Fish, M. Q., Hewer, R., Wallis, C. L., Venter, W. D. F., Stevens, W. S., Papathanasopoulos, M. A., 2010. Natural polymorphisms of integrase among HIV type 1-infected South African patients. AIDS Res. Human Retrovir.; 26: 489-493.

Fitch, W. M., Margoliash, E., 1967. Construction of phylogenetic trees. Science; 20: 279-284.

Freed, E. O., 2001. HIV-1 Replication. Somat. Cell Mol. Genet.; 26: 13-33.

Galai, N., Kalinkovich, A., Burstein, R., Vlahov, D., Bentwich, Z., 1997. African HIV-1 subtype C and rate of progression among Ethiopian immigrants in Israel. Lancet; 349: 180-181.

Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palkerr, T. J., Redfield, R., Oleske, J., Safai, B., White, G., Foster, P., Markham, P. D., 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science; 224: 500-503.

Gao, F., Trask, S. A., Hui, H., Mamaeva, O., Chen, Y., Theodore, T. S., Foley, B. T., Korber, B. T., Shaw, G. M., Hahn, B. H., 2001a. Molecular characterization of a highly divergent HIV type 1 isolate obtained early in the AIDS epidemic from the Democratic Republic of Congo. AIDS Res. Human Retrovir.; 17: 1217-1222.

Gao, F., Vidal, N., Li, Y., Trask, S. A., Chen, Y., Kostrikis, L. G., Ho, D. D., Kim, J., Oh, M.-D., Choe, K., Salminen, M., Robertson, D. L., Shaw, G. M., Hahn, B. H., Peeters, M., 2001b. Evidence of two distinct subsubtypes within the HIV-1 subtype A radiation. AIDS Res. Hum. Retrovir.; 17: 675-688.

Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M.,

Hahn, B. H., 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. Nature; 397: 436-441.

Gao, F., Robertson, D. L., Carruthers, C. D., Morrison, S. G., Jian, B., Chen, Y., Barre-Sinoussi, F., Girard, M., Srinivasan, A., Abimiku, A. G., Shaw, G. M., Sharp, P. M., Hahn, B. H., 1998. A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. J. Virol.; 72: 5680-5698.

Gao, F., Morrison, S. G., Robertson, D. L., Thornton, C. L., Craig, S., Karlsson, G., Sodroski, J., Morgado, M., Galvao-Castro, B., Von Briesen, H., Beddows, S., Weber, J., Sharp, P. M., Shaw, G. M. Hahn, B. H., 1996. Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. The WHO and NIAID Networks for HIV Isolation and Characterization. J. Virol.; 70: 1651-1667.

Gao, F., Yue, L., White, A. T., Pappas, P. G., Barchue, J., Hanson, A. P., Greene, B. M., Sharp, P. M., Shaw, G. M., Hahn, B. H., 1992. Human infection by genetically diverse $SIV_{sm}$-related HIV-2 in West Africa. Nature; 358: 495-499.

Garber, D. A., Silvestri, G., Feinberg, M. B., 2004. Prospects for an AIDS vaccine: three big questions, no easy answers. Lancet Infect. Dis.; 4: 397-413.

Gaschen, B., Taylor, J., Yusim, K., Foley, B., Gao, F., Lang, D., Novitsky, V., Haynes, B., Hahn, B. H., Bhattacharya, T., Korber, B., 2002. Diversity considerations in HIV-1 vaccine selection. Science; 296: 2354-2360.

Gelderblom, H. R., Hausmann, E. H. S., Özel, M., Pauli, G., Koch, M. A., 1987. Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. Virology; 156: 171-176.

Gifford, R. J., Liu, T. F., Rhee, S-Y., Kiuchi, M., Hue, S., Pillay, D., Shafer, R. W., 2009. The calibrated population resistance tool: standardized genotypic

estimation of transmitted HIV-1 drug resistance. Bioinformatics; 25: 1197–1198

Gifford, R. J., De Oliveira, T., Rambaut, A., Pybus, O. G., Dunn, D., Vandamme, A-M., Kellam, P., Pillay, D., on behalf of the UK Collaborative Group on HIV Drug Resistance, 2007. Phylogenetic surveillance of viral genetic diversity and the evolving molecular epidemiology of human immunodeficiency virus type 1. J. Virol.; 81: 13050-13056.

Goldman, N., 1993. Statistical tests of models of DNA substitution. J. Mol. Evol.; 36: 182-198.

Gordon, M., De Oliveira, T., Bishop, K., Coovadia, H. M., Madurai, L., Engelbrecht, S., Janse van Rensburg, E., Mosam, A., Smith, A., Cassol, S., 2003, Molecular characteristics of human immunodeficiency virus type 1 subtype C viruses from KwaZulu-Natal, South Africa: implications for vaccine and antiretroviral control strategies. J. Virol.; 77; 2587-2599.

Gottlieb, M. S., Schroff, R., Schanker, H. M., Weisman, J. D., Fan, P. T., Wolf, R. A., Saxon, A., 1981. *Pneumocystis carinii* pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N. Engl. J. Med.; 305: 1425-1431.

Grossman, Z., Paxinos, E. E., Averbuch, D., Maayan, S., Parkin, N. T., Engelhard, D., Lorber, M., Istomin, V., Shaked, Y., Mendelson, E., Ram, D., Petropoulos, C. J., Schapiro, J. M., 2004a. Mutation D30N is not preferentially selected by human immunodeficiency virus type 1 subtype C in the development of resistance to nelfinavir. Antimicrob. Agents Chemother.; 48: 2159-2165.

Grossman, Z., Istomin, V., Averbuch, D., Lorberc, M., Risenberg, K., Levi, I., Chowers, M., Burke, M., Bar Yaacov, N., Schapiro, J. M., for the Israel AIDS Multi-Center Study Group, 2004b. Genetic variation at NNRTI resistance-associated positions in patients infected with HIV-1 subtype C AIDS; 18: 909-915.

Hahn, B. H., Shaw, G. M., De Cock, K. M. Sharp., P. M., 2000. AIDS as a zoonosis: scientific and public health implications. Science; 287: 607 - 614.

Hall, T. 2005. BioEdit version 7.0.5. Available from http://www.mbio.ncsu.edu/BioEdit/bioedit.html. [Online.]

Hasegawa. M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol.; 21: 160-174.

Hemelaar, J., 2012. The origin and diversity of the HIV-1 pandemic. Trends Mol. Med.; 18: 182-192.

Hemelaar, J., Gouws, E., Ghys, P. D., Osmanov, S., and WHO-UNAIDS Network for HIV isolation and characterization, 2011. Global trends in molecular epidemiology of HIV-1 during 2000–2007. AIDS; 25: 679-689.

Higgins, D. G., Sharp, P. M., 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene; 73: 237-244.

Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M., Markowitz, M., 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature; 373: 123-126.

Holguín, Á., Paxinos, E., Hertogs, K., Womac, C., Soriano, V., 2004. Impact of frequent natural polymorphisms at the protease gene on the *in vitro* susceptibility to protease inhibitors in HIV-1 non-B subtypes. J. Clin. Virol.; 31: 215-220.

Holmes, E. C., Garnett, G. P., 1994. Genes, trees and infections: molecular evidence in epidemiology. Trends Ecol. Evol. 9, 256-260.

Hu, D., Buvé, A., Baggs, J., Van der Groen, G., Dondero, T. J., 1999. What role does HIV-1 subtype play in transmission and pathogenesis? An epidemiological perspective. AIDS; 13: 873-881.

Hu, W. S., Temin, H. M., 1990. Genetic consequences of packing two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. Proc. Natl. Acad. Sci. USA; 87: 1556-1560.

Hu, D. J., Dondero, T. J., Rayfield, M. A., George, R. J., Schochetman, G., Jaffe, H. W., Luo, C-C., Kalish, M. L., Weniger, B. G., Pau, C-P., Schable, C. A., Curran, J. W., 1996. The emerging genetic diversity of HIV, the importance of global surveillance for diagnostics, research, and prevention. JAMA. 275: 210-216.

Hu, W-S., Temin, H. M., 1990. Retroviral recombination and reverse transcription. Science; 250: 1227-1233.

Huang, K-H. G., Goedhals, D., Fryer, H., Van Vuuren, C., Katzourakis, A., De Oliveira, T., Brown, H., Cassol, S., Seebregts, C., McLean, A., Klenerman, P., Phillips, R., Frater, J., and the Bloemfontein-Oxford Collaborative Group, 2009. Prevalence of HIV type-1 drug-associated mutations in pre-therapy patients in the Free State, South Africa. Antivir. Ther.; 14: 975-984.

Huang, W., Eshleman, S. E., Toma, J., Fransen, S., Stawiski, E., Paxinos, E., Whitcomb, J. M., Young, A. M., Donnell, D., Mmiro, F., Musoke, P., Guay, L. A., Jackson, J. B., Parkin, N. T., Petropoulos, C. J., 2007. Coreceptor tropism in human immunodeficiency virus type 1 subtype D: high prevalence of CXCR4 tropism and heterogeneous composition of viral populations. J. Virol., 81: 7885-7893.

Hunt, G. M., Ledwaba, J., Basson, A. E., Moyes, J., Cohen, C., Singh, B., Bertagnolio, S., Jordan, M. R., A. Puren, A., Morris, L., 2012. Surveillance of transmitted HIV-1 drug resistance in Gauteng and KwaZulu-Natal Provinces, South Africa, 2005–2009. Clin. Infect. Dis.; 54: S334-S338.

Hunt, G.M., Johnson, D., Tiemessen, C.T., 2001. Characterisation of the long terminal repeat regions of South African human immunodeficiency virus type 1 isolates. Virus Genes; 23: 27-34.

Hunt, G., Tiemessen, C. T., 2000. Occurrence of additional NF-κB-binding motifs in the long terminal repeat region of South African HIV type 1 subtype C Isolates. AIDS Res. Human Retrovir.; 16: 305-306.

Iweriebor, B. C., Mavhandu, L. G., Masebe, T., Rekosh, D., Hammarskjold, M-L., Mphahlele, J., Bessong, P. O., 2012. Molecular epidemiology of HIV in two highly endemic areas of northeastern South Africa. Arch. Virol.; 157: 455-465.

Jacobs, G. B., Loxton, A. G., Laten, A., Robson, B., Janse van Rensburg, E., Engelbrecht, S., 2009. Emergence and diversity of different HIV-1 subtypes in South Africa, 2000-2001. J. Med. Virol.; 81: 1852-1859.

Jacobs, G.B., Nistal, M., Laten, A., Janse van Rensburg, E., Rethwilm, A., Preiser, W., Bodem, J., Engelbrecht, S., 2008a. Molecular analysis of HIV type 1 *vif* sequences from Cape Town, South Africa. AIDS Res. Human Retrovir. 24, 991-994.

Jacobs, G. B., A. Laten, E. J. van Rensburg, J. Bodem, B. Weissbrich, A. Rethwilm, W. Preiser, Engelbrecht, S., 2008b. Phylogenetic diversity and low level antiretroviral resistance mutations in HIV type 1 treatment-naïve patients from Cape Town, South Africa. AIDS Res. Human Retrovir. 24: 1009-1012.

Jacobs, G.B., Loxton, A.G., Laten, A., Engelbrecht, S., 2007. Complete genome sequencing of a non-syncytium-inducingHIV type 1 subtype D strain from Cape Town, South Africa. AIDS Res. Human Retrovir. 23, 1575-1578.

Jacobs, G. B., De Beer, C., Fincham, J. E., Adams, V., Dhansay, M. A., Janse van Rensburg, E., and S. Engelbrecht, S., 2006. Serotyping and genotyping of HIV-1 infection in residents of Khayelitsha, Cape Town, South Africa. J.Med. Virol.; 78: 1529-1536.

Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, Jr, A. D., Lu X., Tantillo, C., Williams, R., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes S., Arnold E., 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. Proc. Natl. Acad. Sci. USA; 90: 6320-6324.

Janssens, W., Buvé, A., Nkengasong, J. N., 1997. The puzzle of HIV-1 subtypes in Africa. AIDS; 11: 705-712.

Jetzt, A. E., Yu, H., Klarmann, G. J., Ron. Y., Preston, B. D., Dougherty, J., 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome. J. Virol.; 74: 1234-1240.

Johnson, V. A., Brun-Vézinet, F., Clotet, B., Günthard, H. F., Kuritzkes, D. R., Pillay, D., Schapiro, J. M., Richman, D. D., 2010. Update of the drug resistance mutations in HIV-1: December 2010. Top. HIV Med.; 18:156-163.

Jülg, B., Goebel, F.D., 2005. HIV Genetic Diversity: Any Implications for Drug Resistance? Infection; 33: 299-301.

Kantor, R., Katzenstein, D. A., Efron, B., Carvalho, A. P., Wynhoven, B., Cane, P., Clarke, J., Sirivichayakul, S., Soares, M. A., Snoeck, J., Pillay, C., Rudich, H., Rodrigues, R., Holguin, A., Ariyoshi, K., Bouzas, M. B., Cahn, P., Sugiura, W., Soriano, V., Brigido, L. F., Grossman, Z., Morris, L., Vandamme, A-M., Tanuri, A., Phanuphak, P., Weber, J. N., Pillay, D., Harrigan, P. R., Camacho, R., Schapiro, J. M., Shafer, R. W., 2005. Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration. Plos Med.; 2: 0325-0337.

Kawashima, Y., Pfafferott, K., Frater, J., Matthews, P., Payne, R., Addo, M., Gatanaga,H., Fujiwara, M., Hachiya, A., Koizumi, H., Kuse, N., Oka, S., Duda, A., Prendergast, A., Crawford, H., Leslie, A., Brumme, Z., Brumme, C., Allen, T., Brander, C., Kaslow, R., Tang, J., Hunter, E., Allen, S., Mulenga, J., Branch, S., Roach, T., John, M., Mallal, S., Ogwu, A., Shapiro, R., Prado, J. G., Fidler, S., Weber, J., Pybus, O. G., Klenerman, P., Ndung'u, T., Phillips, R., Heckerman, D., Harrigan, P. R., Walker, B. D., Takiguchi, M., Goulder, P., 2009. Adaptation of HIV-1 to human leukocyte antigen class I. Nature 458: 641-646.

Keele, B. F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M. L., Bibollet-Ruche, F., Chen, Y., Wain, L. V., Liegeois, F., Loul, S., Mpoudi Ngole, E., Bienvenue, Y., Delaporte, E., Brookfield, J. F. Y., Sharp, P. M., Shaw, G. M., Peeters, M., Hahn, B. H., 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science; 313: 523-526.

Kimura, M., 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol.; 16: 111-120.

Kiwanuka, N., Laeyendecker, O., Quinn, T. C., Wawer, M. J., Shepherd, J., Robb, M., Kigozi, G., Kagaayi, J., Serwadda, D., Makumbi, F. E., Reynolds, S. J., Gray, R. H., 2009. HIV-1 subtypes and differences in heterosexual HIV transmission among HIV-discordant couples in Rakai, Uganda. AIDS; 23: 2479-2484.

Kiwanuka, N., Laeyendecker, O., Robb, M., Kigozi, G., Arroyo, M., McCutchan F., Eller, L. A., Eller, M., Makumbi, F., Birx, D., Wabwire-Mangen, F., Serwadda, D., Sewankambo, N. K., Quinn, T. C., Wawer, M., Gray, R., 2008. Effect of human immunodeficiency virus type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection. J. Infect. Dis.; 197: 707-713.

Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., Steitz, T. A., 1992. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science; 256: 1783-1790.

Kosakovsky Pond, S. L., Posada, D., Stawiski, E., Chappey, C., Poon, A. F.Y., Hughes, G., Fearnhill, E., Gravenor, M. B., Leigh Brown, A. J., Frost, S. D.W., 2009. An evolutionary model-based algorithm for accurate phylogenetic breakpoint mapping and subtype prediction in HIV-1. PLoS Comput. Biol.; 5: e1000581.

Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., Detours, V., 2001. Evolutionary and immunological implications of contemporary HIV-1 variation. British Med. Bull.; 58: 19-42.

Korber, B. T., Foley, B. T., Kuiken, C. L., Pillai, S. K., Sodroski J. G., 1998. Numbering positions in HIV relative to HXB2CG. 102-111. (Los Alamos National Laboratory; http://www.hiv.lanl.gov).

Kostrikis, L. G., Cao, Y., Ngai, H., Moore, J. P., Ho., D. D., 1996. Quantitative analysis of serum neutralization of human immunodeficiency virus type 1 from subtypes A, B, C, D, E, F, and I: lack of direct correlation between neutralization serotypes and genetic subtypes and evidence for prevalent serum-dependent infectivity enhancement. J. Virol.; 70: 445-458.

Koulinska, I. N., Ndung'u, T., Mwakagile, D., Msamanga, G., Kagoma, C., Fawzi, W., Essex, M., Renjifo, B., 2001. A new human immunodeficiency virus type 1 circulating recombinant form from Tanzania. AIDS Res. Hum. Retrovir.; 17: 423-431.

Kuiken, C., Korber, B., 1998. Sequence Quality Control. pp. III-80-90 in Human Retroviruses and AIDS 1998. Edited by: Korber B, Kuiken CL, Foley B, Hahn B, McCutchan F, Mellors JW, and Sodroski J. Published by: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

Kuiken, C., Leitner, T., Foley, B., Hahn, B., Marx, P., McCutchan, F., Wolinsky, S., Korber, B., (Editors), 2008. HIV Sequence Compendium 2008. Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico.

Laukkanen, T., Albert, J., Liitsola, K., Green, S. D., Carr, J. K., Leitner, T., McCutchan, F. E., Salminen, M. O., 1999. Virtually full-length sequences of HIV type 1 subtype J reference strains. AIDS Res. Hum. Retrovir.; 15: 293-297.

Learn, G. H., Mullins, J. I., 2003. The Microbial Forensic Use of HIV Sequences. pp. 22-37 in HIV Sequence Compendium 2003. Edited by: Leitner T, Foley B, Hahn B, Marx P, McCutchan F, Mellors J, Wolinsky S, and Korber B, editors. Published by Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, LA-UR number 04-7420.

Leitner, T., Kumar, S., Albert, J., 1997. Tempo and mode of the nucleotide substitution in *gag* and *env* gene fragments in human immunodeficiency virus type 1 populations with a known transmission history. J. Virol.; 71: 4761-4770.

Leitner, T., Korber, K., Robertson, D., Gao, F., Hahn, B., 1997. Updated Proposal of Reference Sequences of HIV-1 Genetic Subtypes. pp. III-19-24 in Human Retroviruses and AIDS 1997. Edited by: Korber B, Hahn B, Foley B, Mellors JW, Leitner T, Myers G, McCutchan F and Kuiken CL. Published by: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

Létourneau, S., Im, E-J., Mashishi, T., Brereton C., Bridgeman A., Yang H., Dorrell, L., Dong, T., Korber, B., McMichael, A. J., Hanke, T., 2007. Design and pre-clinical evaluation of a universal HIV-1 vaccine. PLoS ONE 2 (10): e984.

Lian, Y., Srivastava, I., Gómez-Román, V. R., Zur Megede, J., Sun, Y., Kan, E., Hilt, S. Engelbrecht, S., Himathongkham, S., Luciw, P. A., Otten, G., Ulmer, J. B., Donnelly, J. J., Rabussay, D., Montefiori, D., Janse van Rensburg, E., Barnett, S. W., 2005. Evaluation of envelope vaccines derived from the South African subtype C human immunodeficiency virus type 1 TV 1 strain. J. Virol.; 79: 13338-13349.

Lihana, R. W., Ssemwanga, D., Abimiku, A., Ndembi, N., 2012. Update on HIV-1 diversity in Africa: a decade in review. AIDS Rev.; 14: 83-100.

Lindström, A., Albert, J., 2003. A simple and sensitive ‘in-house’ method for determining genotypic drug resistance in HIV-1. J. Virol. Methods; 107: 45-51.

Liò, P., Goldman, N., 1998. Models of molecular evolution and phylogeny. Genome Res.; 8: 1233-1244.

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W. Ray, S. C. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol.; 73: 152-160.

Louwagie, J., McCuthan, F. E., Peeters, M., Brennan, T. P., Sanders, B. E., Eddy, G. A., Van der Groen, G., Fransen, K., Gershy-Damet, G.-M., Deleys,

R., Burke, D. S., 1993. Phylogeneitc analysis of *gag* genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. AIDS; 7: 769-780.

Loxton, A. G., Treurnicht, F., Laten, A., Janse van Rensburg, E., Engelbrecht, S., 2005. Sequence analysis of near full-length HIV type 1 subtype D primary strains isolated in Cape Town, South Africa, from 1984 to 1986. AIDS Res. Human Retrovir.; 21: 410-413.

Lurie, M. N., Williams, B. G., Zuma, K., Mkaya-Mwamburi, D., Garnett, G. P., Sweat, M. D., Gittelsohn, J., Abdool Karim, S. S., 2003. Who infects whom? HIV-1 concordance and discordance among migrant and non-migrant couples in South Africa. AIDS; 17: 2245-2252.

Lurie, M. 2000. Migration and AIDS in southern Africa: a review. South African Journal of Science; 96: 343-347.

Makuwa, M., Souquière, S., Niangui, M. T., Rouquet, P., Apterei, C., Roques, P., Simon, F., 2002. Reliability of rapid diagnostic tests for HIV variant infection. J Virol. Methods; 103: 183-190.

Malim, M. H., Emerman, M., 2001. HIV-1 Sequence variation: drift, shift, and attenuation. Cell; 104: 469-472.

Mansky, L. M., Temin, H. M., 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J. Virol.; 69: 5087-5094.

Marconi, V. C., Sunpath, H., Lu, Z., Gordon, M., Koranteng-Apeagyei, K., Hampton, J., Carpenter, S., Giddy, J., Ross, D., Holst, H., Losina, E., Walker, B. D., Kuritzkes, D. R., for the South Africa Resistance Cohort Study Team, 2008. Prevalence of HIV-1 drug resistance after failure of a first highly active antiretroviral therapy regimen in KwaZulu Natal, South Africa. Clin. Infect. Dis.; 46: 1589-1597.

McCormick-Davis, C., Dalton, S. B., Singh, D. K., Stephens, E. B., 2000. Comparisons of Vpu sequences from diverse geographical isolates of HIV

type 1 identifies the presence of highly variable domains, additional invariant amino acids and a signature sequence motif common to subtype C isolates. AIDS Res. Hum. Retrovir.; 16: 1089-1095.

McCutchan, F. E. 1999. Global diversity in HIV. In The Evolution of HIV. Edit, Crandall, K. A., The John Hopkins Unversity Press, Baltimore and London.

Metzker, M. L., Mindell, D. P., Liu, X. M., Ptak, R. G., Gibbs, R. A., Hillis, D. M., 2002. Molecular evidence of HIV- 1 transmission in a criminal case. Proc. Natl. Acad. Sci. U S A, 99: 14292-14297.

Ministry of Health, Mozambique, 2005. Report on the update of the HIV epidemiological surveillance data — 2004 Round, August. National Directorate of Health, Ministry of Health, Maputo.

Montes, B., Vergne, L., Peeters, M., Reynes, J., Delaporte, E., Segondy, M., 2004. Comparison of drug resistance mutations and their interpretation in patients infected with non-B HIV-1 variants and matched patients infected with HIV-1 subtype B. J. Acquir. Immune Defic. Syndr.;35: 329-336.

Moore, J. P., Parren, P. W. H. I., Burton, D. R., 2001. Genetic subtypes, humoral immunity, and human immunodeficiency virus type 1 vaccine development. J. Virol.; 75: 5721-5729.

Moore, J. P., Cao, Y., Leu, J., Korber, B., Ho, D. D., 1996. Inter- and intraclade neutralization of human immunodeficiency virus type 1: genetic clades do not correspond to neutralization serotypes but partially correspond to gp120 antigenic serotypes. J. Virol., 70: 427-444.

Morris, L., Pillay, C., Chezzi, C., Lupondwana, P., Ntsala, M., Levin, L., Venter, F., Martinson, N., Gray, G., McIntyre, J., 2003. Low frequency of the V106M mutation among HIV-1 subtype C-infected pregnant women exposed to nevirapine. AIDS; 17: 1698-1700.

Neilson, J. R., John, G. C., Carr, J. K., Lewis, P., Kreiss, J. K., Jackson, S., Nduati, R. W., Mbori-Ngacha, D., Panteleeff, D. D., Bodrug, S., Giachetti, C., Bott, M. A., Richardson, B. A., Bwayo, J., Ndinya-Achola, J., Overbaugh, J.

1999. Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya. J. Virol.; 73: 4393-4403.

Nkengasong, J. N., Janssens, W., Heyndrickx, L., Fransen, K., Ndumbe, P. M., Motte, J., Leonaers, A., Ngolle, M., Ayuk, J., Piot, P., Van der Groen, G., 1994. Genotypic subtypes of HIV-1 in Cameroon. AIDS. 8: 1405-1412.

Nora, T., Charpentier, C., Tenaillon, O., Hoede, C., Clavel, F., Hance, A. J., 2007. Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. J. Virol.; 81: 7620-7628

Novitsky, V., Wester, C. W., Degruttola, V., Bussmann, H., Gaseitsiwe, S., Thomas, A., Moyo, S., Musonda, R., Van Widenfelt, E., Marlink, R. G., Essex, M., 2007. The reverse transcriptase 67N 70R 215Y genotype is the predominant TAM pathway associated with virologic failure among HIV Type 1C-infected adults treated with ZDV/ddI-containing HAART in Southern Africa. AIDS Res. Human Retrovir.; 23: 868-878.

Novitsky, V., Smith, U. R., Gilbert, P., McLane, M. F., Chigwedere, P., Williamson, C., Ndung'u, T., Klein, I., Chang, S.-Y., Peter, T., Thior, I., Foley, B. T., Gaolekwe, S., Rybak, N., Gaseitsiwe, S., Vannberg, F., Marlink, R., Lee, T.-H., Essex, M., 2002. Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design? J. Virol.; 76: 5435-5451.

Novitsky, V. A., Montano, M. A., McLane, M. F., Renjifo, B., Vannberg, F., Foley, B. T., Ndung'u, T. P., Rahman, M., Makhema, M. J., Marlink, R., Eessex, M., 1999. Molecular cloning and phylogenetic analysis of human immunodeficiency virus type 1 subtype C: a set of 23 full-length clones from Botswana. J. Virol., 73: 4427-4432.

Overbaugh, J., Morris, L., 2012. Antibody response against HIV-1. Cold Spring Harb. Perspect. Med.; 2: a007039.

Palmer, S., Alaeus, A., Albert, J., Cox, S., 1998. Drug susceptibility of subtypes A, B, C, D, and E HIV-1 primary isolates. AIDS Res. Hum. Retrovir.; 14: 157-162.

Papathanasopoulos, M. A., Meyers, T. P. T. M. A., McCutchan, F. E., Morris, L., 2003. Full-length genome characterization of HIV type 1 subtype C isolates from two slow-progressing perinatally infected siblings in South Africa. AIDS Res. AND Human Retrovir.; 19: 1033-1037.

Papathanasopoulos, M. A., Cilliers, T., Morris, L., Mokili, J. L., Dowling, W., Brix, D. L., McCutchan, F. E., 2002. Full-length genome analysis of HIV-1 subtype C utilizing CXCR4 and intersubtype recombinants isolated in South Africa. AIDS Res. Human Retrovir.; 18, 879-886.

Paraskevis, D., Hatzakis, A., 1999. Molecular epidemiology of HIV-1 infection. AIDS Rev; 1: 238-249.

Peeters, M., Vincent, R., Perret, J. L., Lasky, M., Patrel, D., Liegeois, F., Courgnaud, V., Seng, R., Matton, T., Molinier, S., Delaporte, E., 1999. Evidence for differences in MT2 cell tropism according to genetic subtypes of HIV-1: syncytium-inducing variants seem rare among subtype C HIV-1 viruses. J. Acq. Imm. Def. Syn. Human Retro. 20: 115-121.

Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M., Ho, D. D., 1996. HIV-1 dynamics *in vivo*: virion clearance rate, infected cell life-span, and viral generation time. Science; 271: 1582-1586.

Perrin, L., Kaiser, L., Yerly, S., 2003. Travel and the spread of HIV-1 genetic variants. Lancet Infect. Dis.; 3: 22-27.

Pillay, D., Sinka, K., Rice, P., Peters, B., Clarke. J., Workman, J., Evans, B., Cane, P.A., 2000. Impact of HIV-1 subtype on NNRT resistance mutations. Antivir. Ther.; (Suppl 3): 128, Abstract 163.

Ping, L. H., Nelson, J. A., Hoffman, I. F., Schock, J., Lamers, S. L., Goodman, M., Vernazza, P., Kazembe, P., Maida, M., Zimba, D., Goodenow, M. M., Eron, Jr., J. J., Fiscus, S. A., Cohen, M. S., Swanstrom. R., 1999.

Characterization of V3 sequence heterogeneity in subtype C human immunodeficiency virus type 1 isolates from Malawi: underrepresentation of X4 variants. J. Virol. 73:6271-6281.

Plantier, J-C., Leoz, M., Dickerson, J. E., De Oliveira, F., Cordonnier, F., Lemée, V., Damond, F., Robertson, D. L., Simon, F., 2009. A new human immunodeficiency virus derived from gorillas. Nat. Med.; 15: 871-872.

Plantier, J-C., Dachraoui, R., Lemée, V., Gueudin, M., Borsa-Lebas, F., Caron, F., Simon, F., 2005. HIV-1 resistance genotyping on dried serum spots. AIDS; 19: 391-397.

Powell, R. L., Zhao, J., Konings, F. A., Tang, S., Nanfack, A., Burda, S., Urbanski, M. M., Saa, D. R., Hewlett, I., Nyambi, P. N., 2007. Identification of a novel circulating recombinant form (CRF) 36_cpx in Cameroon that combines two CRFs (01_AE and 02_AG) with ancestral lineages of subtypes A and G. AIDS Res. Hum. Retrovir.; 23: 1008-1019.

Preston, B. D., Poiesz, B. J., Loeb, L. A., 1988. Fidelity of HIV-1 reverse transcriptase. Science; 242: 1169-1171.

Preston, B. D., Dougherty, J. P., 1996. Mechanisms of retroviral mutation Trends in Microbiol. 4: 16-21.

Rambaut, A., Posada, D., Crandall, K. A., Holmes, E. C., 2004.The causes and consequences of HIV evolution. Nature Rev. Genetics; 5: 52-61.

Rambaut, A., Robertson, D. L., Pybus, O. G., Peeters, M., Holmes, E. C., 2001. Phylogeny and the origin of HIV-1. Nature; 410: 1047-1048.

Ramírez de Arellano, E., Soriano, V., Alcami, J., Holguín, Á., 2006. New findings on transcription regulation across different HIV-1 subtypes. AIDS Rev.; 8: 9-16.

Ramjee, G., Gouws, E., 2002. Prevalence of HIV among truck drivers visiting sex workers in KwaZulu-Natal, South Africa. Sex Trans. Dis.; 29: 44-49.

Renjifo, B., Gilbert, P., Chaplin, B., Msamanga, G., Mwakagile, D., Fawzi, W., Tanzanian Vitamin and HIV Study Group, Essex, M., 2004. Preferential *in-utero* transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D. AIDS; 18: 1629-1636.

Rhee, S-Y., Gonzales, M. J., Kantor, R., Betts, B. J., Ravela, J., Shafer, R. W., 2003. Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res.; 31: 298-303.

Robertson, D, L., Anderson, J., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P. M., Wolinsky, S., Korber, B., 2000. HIV- 1 nomenclature proposal. Science; 288: 55-56.

Robertson, D. L., Sharp, P. M., McCutchan, F. E., Hahn, B. H., 1995. Recombination in HIV-1. Nature; 374: 124-126.

Rodenburg, C. M., Li, Y., Trask, S. A., 2001. Near full-length clones and reference sequences for subtype C isolates of HIV type 1 from three different continents. AIDS Res. Hum. Retrovir., 17: 161-168.

Romani, B., Glashoff, R., Engelbrecht, S., 2009. Molecular and phylogenetic analysis of HIV type 1 *vpr* sequences of South African strains. AIDS Res Human Retrovir.; 25: 357-362.

Rousseau, C. M., Birditt, B. A., McKay, A. R., Stoddard, J. N., Lee, T. C., McLaughlin, S., Moore S. W., Shindo, N., Learn, G. H., Korber, B. T., Brander, C., Goulder, P. J. R., Kiepiela, P., Walker, B. D., Mullins, J. I., 2006. Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. J. Virol. Methods 136: 118-125.

Rubbert, A., Behrens, G., Ostrowski, M., 2007. Pathogenesis of HIV-1 infection. In Hoffmann, C., Rockstroh, J. K., Kamps, B. S. (editors) HIV Medicine 2007. Flying Publisher. Paris, Cagliari, Wuppertal, 15th Edition; 59-86.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol.; 4: 406-425.

Salminen, M. O., Johansson, B., Sonnerborg, A., Ayehunie, S., Gotte, D., Leinikki, P., Burke, D. S., McCutchan, F. E., 1996. Full-length sequence of an Ethiopian human immunodeficiency virus type 1 (HIV-1) isolate of genetic subtype C. AIDS Res. Hum. Retrovir.; 12: 1329-1339.

Sanabani, S., Neto, W. K., De Sa Filho, D. J., Diaz, R. S., Munerato, P., Janini, L. M. Sabino, E. C. 2006. Full-length genome analysis of human immunodeficiency virus type 1 subtype C in Brazil. AIDS Res. Hum. Retrovir.; 22: 171-176.

San Mauro, D., Agorreta, A., 2010. Molecular sytematics: a synthesis of the common and the state of knowledge. Cell Mol. Biol. Lett.; 15: 311-341.

Sarafianos, S. G., Das, K., Hughes, S. H., Arnold, E., 2004. Taking aim at a moving target: designing drugs to inhibit drug-resistant HIV-1 reverse transcriptases. Curr. Opin. Struct. Biol.; 14: 716-730.

Sarafianos, S. G., Kalyan Das, K., Tantillo, C., Clark,Jr A. D., Ding, J., Whitcomb, J. M., Boyer, P. L., Hughes, S. H., Arnold, E., 2001. Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA: DNA. EMBO J. 20: 1449-1461.

Schultz, A.-K. Zhang, M. Bulla, I. Leitner, T. Korber, B. Morgenstern, B., Stanke, M., 2009. jpHMM: Improving the reliability of recombinantion prediction in HIV-1. Nucleic Acids Res.; 37: 647-651.

Schultz, A.-K. Zhang, M. Leitner, T. Kuiken, C. Korber, B. Morgenstern, B. Stanke M., 2006. A Jumping Profile Hidden Markov Model and apllication to recombination sites in HIV and HCV genomes BMC Bioinformatics; 7: 265.

Shafer, R. W., Rhee, S-Y., Bennett, D. E., 2008. Consensus drug resistance for epidemiological surveillance: basic principles and potential controversies. Antivir. Ther.; 13 (Supp2): 59-68.

Shafer, R., Chuang, T., Hsu, P., White, C., Katzenstein, D., 1999. Sequence and drug susceptibility of subtype C protease from HIV-1 seroconverters in Zimbabwe. AIDS Res. Hum. Retrovir.; 15: 65-69.

Shankarappa, R., Chatterjee, R., Learn, G. H, Neogi, D., Ding, M., Roy, P., et al. 2001. Human immunodeficiency virus type 1 *env* sequences from Calcutta in Eastern India: identification of features that distinguish subtype C sequences in India from other subtype C sequences. J. Virol.; 75: 10479-10487.

Sharp, P. M., 2002. Origins of human virus diversity. Cell; 108: 305-312.

Sharp, P. M., Bailes, E., Robertson, D. L., Gao, F., Hahn, B. H., 1999. Origins and evolution of Aids viruses. Biol. Bull.; 196: 338-342.

Shott, J P., Galiwango, R. M., Reynolds, S. J., 2012. A quality management approach to implementing point-of-care technologies for HIV diagnosis and monitoring in sub-Saharan Africa. J. Trop. Med.; 651927. doi: 10.1155/2012/651927.

Simon, F., Mauclere, P., Roques, P., Loussert-Ajaka,I., Muller-Trutwin, M. C., Saragosti, S., Georges-Courbot, M. C., Barre-Sinoussi, F., Brun-Vezinet, F., 1998. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nat. Med.; 4: 1032-1037.

Sluis-Cremer, N., Kempner, E., Simon, F., Mauclere, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M. C., Saragosti, S., Georges-Courbot, M. C., Barre-Sinoussi, F., Brun-Vezinet, F., 1998. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nat. Med.; 4: 1032-1037.

Sizmann, D., Glaubitz, J., Simon, C. O., Goedel, S., Buergisser, P., Drogan, D., Hesse, M., Kröh, M., Simmler, P., Dewald, M., Gilsdorf, M., Fuerst, M., Ineichen, R., Kirn, A., Pasche, P., Wang, Z., Weisshaar, S., Young, K., Haberhausen, G., Babiel, R., 2010. Improved HIV-1 RNA quantitation by

COBAS® AmpliPrep/COBAS® TaqMan® HIV-1 Test, v2.0 using a novel dual-target approach. J. Clin. Virol.; 49: 41-46.

Skar, H., Hedskog, C., Albert, J., 2011. HIV-1 evolution in relation to molecular epidemiology and antiretroviral resistance. Ann. N.Y. Acad. Sci.; 1230: 108-118.

Sluis-Cremer, N., D., Arion, D., Parniak, M. A., 2000. Molecular mechanisms of HIV-1 resistance to nucleoside reverse transcriptase inhibitors (NRTIs). Cell. Mol. Life Sci.; 57: 1408-1422.

Soares, E. A., Santos, A. F., Gonzalez, L. M., Lalonde, M. S., Denis M. Tebit, D. M., Tanuri, A., Arts, E. J., Soares, M. A., 2009. Mutation T74S in HIV-1 subtype B and C proteases resensitizes them to ritonavir and indinavir and confers fitness advantage. J. Antimicrob. Chemother.; 64: 938-944.

South African Government, 2012. Global AIDS Response Progress Report 2012. Pretoria, South Africa. (http://www.doh.gov.za/list.php?type=HIV%20and%20AIDS) (Accessed: 14 August 2012).

South African Government, 2010. National Department of Health. The National Antenatal Sentinel HIV and Syphilis Prevalence Survey, South Africa, 2010. National Department of Health, Pretoria, South Africa. (http://www.doh.gov.za/list.php?type=HIV%20and%20AIDS) (Accessed: 18 June 2012).

South African Government, 2008. National Department of Health UNGASS Country Progress Reports, South Africa, Pretoria. (http://www.doh.gov.za/list.php?type=HIV%20and%20AIDS) (Accessed: 18 June 2012).

South African Government, 2007. HIV and AIDS and STI strategic plan for South Africa, 2007-2011, South Africa, Pretoria. (http://www.doh.gov.za/list.php?type=HIV%20and%20AIDS) (Accessed: 18 June 2012).

South African Government, 2005. Department of Provincial and Local Government, Bushbuckridge Nodal Economic Development Profile, Mpumalanga, South Africa. (http://www.cogta.gov.za/index.php/documents.html) (Accessed: 16 April 2005).

Sundquist, W. I., Kräusslich, H-G., 2012. HIV-1 Assembly, budding, and maturation. Cold Spring Harb. Perspect. Med.; 2: a006924.

Swanson, P., Devare, S. G., Hackett, Jr., J., JR., 2003. Molecular characterization of 39 HIV-1 isolates representing group M (subtypes A–G) and group O: sequence analysis of *gag* p24, *pol* integrase, and *env* gp41. AIDS Res. Hum. Retrovir; 19: 625-629.

Takehisa, J., Kraus, M. H., Ayouba, A., Bailes, E., Van Heuverswyn, F., Decker, J. M., Li, Y., Rudicell, R. S., Learn, G. H., Neel, C., Mpoudi Ngole, E., Shaw, G. M., Peeters, M., Sharp, P. M., Hahn, B. H., 2009. Origins and biology of simian immunodeficiency virus in wild-living western gorillas. J. Virol.; 83: 1635-1648.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony methods. Mol. Biol. Evol. 28: 2731-2739.

Tatsumi, M., Hara, T., Sakamoto, Y., Terunuma, H., Ichiyama, H., Kasolo, M., Honda, M., Yamamoto, N. Construction and characterization of HIV-1 subtype C infectious molecular clones derived from Zambian isolates. Unpublished.

Tatt, I. D., Barlow, K. L., Clewley, J. P., Gill, O. N., Parry, J. V., 2004. Surveillance of HIV-1 subtypes among heterosexuals in England and Wales, 1997-2000. J. Acquir. Immune Defic. Syndr.; 36: 1092-1099.

Tatt, I. D., Barlow, K. L., Nicoll, A., Clewley, J. P., 2001. The public health significance of HIV-1 subtypes. AIDS; 15: S59-S71.

Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol.; 10: 512-526.

Tatt, I. D., Barlow, K. L., Nicoll, A., Clewley, J. P., 2001. The public health significance of HIV-1 subtypes. AIDS; 15: S59-S71.

Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences, In Lectures on Mathematics in the Life Sciences, (Miura, R. M., ed.) Amer. Math. Soc., Providence, USA.

Taylor, B. S. Sobieszczyk, M. E. McCutchan, F. E. Hammer, S. M. 2008. The challenge of HIV-1 subtype diversity. N. Engl. J. Med.; 358: 1590-1602.

Temin, H. M., 1993. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. Proc. Natl. Acad. Sci. USA.; 90: 6900-6903.

Tscherning, C., Alaeus, A., Fredriksson, R., Björndal, Å., Deng, H., Littman, D. R., Fenyö, E. M., 1998. Differences in chemokine coreceptor usage between genetic subtypes of HIV- 1. Virology, 241: 181-188.

Thompson, J. D., Plewniak, F., Poch, O., 1999. A comprehensive comparison of multiple sequence alignment programs. Nucleic Acids Res.; 27: 2682-2690.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., Higgins, D. G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res.; 24: 4876-4882.

Treurnicht, F.K., Smith, T-L., Engelbrecht, S., Claassen, M., Robson, B.A., Zeier, M., Van Rensburg, E.J., 2002. Genotypic and phenotypic analysis of the *env* gene from South African HIV-1 subtype B and C isolates. J. Med. Virol.; 68: 141-146.

Triques, K., Bourgeois, A., Vidal, N., Mpoudi-Ngole, E., Mulanga-Kabeya, C., Nzilambi, N., Torimiro, N., Saman, E., Delaporte, E., Peeters, M., 2000. Near-

Full-Length Genome Sequencing of Divergent African HIV Type 1 Subtype F Viruses Leads to the Identification of a New HIV Type 1 Subtype Designated K. AIDS Res. Hum. Retrovir; 16: 139-151.

Turner, B. G., Summers, M. F., 1999. Structural biology of HIV. J. Mol. Biol.; 285: 1-32.

UNAIDS, 2012. Report on the global Aids epidemic Geneva UNAIDS)

(http://www.unaids.org/en/KnowledgeCentre/HIVData/GlobalReport/2008/2008_Global_report.asp)

Vallari, A., Holzmayer, V., Harris, B., Yamaguchi, J., Ngansop, C., Makamche, F., Mbanya, D., Kaptué, L., Ndembi, N., Gürtler, L., Devare, S., Brennan, C. A., 2011. Confirmation of putative HIV-1 group P in Cameroon. J. Virol., 85: 1403-1407.

Van Beveren, C. P., Coffin, J., Hughes, S., 1985. Appendix B: HTLV-3/LAV genome. (In) Weiss, R. L., Teich, N., Varmus, H., Coffin, J. (Eds.); RNA Tumor Viruses, Second Edition, 2, Vol. 2: 1102-1123; Cold Spring Harbor Laboratory, Cold Spring Harbor.

Vandamme, A-M., Basic concepts of molecular evolution. In The Phylogenetic Handbook, A Practical Approach To DNA and Protein Phylogeny. Edit: Salemi, M., Vandamme, A-M., Cambridge University Press. United Kingdom. 2003.

Van Griensven, F., 2007. Men who have sex with men and their HIV epidemics in Africa. AIDS; 21: 1361-1362.

Van Harmelen, J., Williamson, C., Kim, B., Morris, L., Carr, J., Karim, S.S., McCutchan, F., 2001. Characterization of full-length HIV type 1 subtype C sequences from South Africa. AIDS Research and Human Retroviruses 17, 1527-1531.

Van Harmelen, J. H., Van der Ryst, E., Loubser, A. S., York, D., Madurai, S., et al. 1999. A predominantly HIV-1 subtype C-restricted epidemic in South African urban populations. AIDS Res. Hum. Retrovir.; 15: 395-398.

Van Harmelen, J., Wood, R., Lambrick, M., Rybicki, E.P., Williamson, A.L., Williamson, C., 1997. An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa. AIDS 11, 81-87.

Van Heuverswyn, F., Yingying Li, Y., Neel, C., Bailes, E., Keele, B. F., Liu, W., Loul, S., Butel, C., Liegeois, F., Bienvenue, Y., Mpoudi Ngolle, E., Sharp, P. M., Shaw, G. M., Delaporte, E., Hahn, B. H., Peeters, M., 2006. Human immunodeficiency viruses: SIV infection in wild gorillas. Nature; 444: 164.

Vartanian, J.-P., Meyerhans, A., Sala, M., Wain-Hobson, S., 1994. G to A hypermutation of the human immunodeficiency virus type 1 genome: evidence for dCTP pool imbalance during reverse transcription. Proc. Natl. Acad. Sci. USA; 91: 3092-3096.

Vasan, A., Renjifo, B., Hertzmark, E., Chaplin, B., Msamanga, G., Essex, M., Fawzi, W., Hunter, D., 2006. Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. Clin. Infect. Dis.; 42: 843-852.

Velazquez-Campoy, A., Todd, M. J., Vega, S., Freire, E., 2001. Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. Proc. Natl. Acad. Sci. USA; 98: 6062-6067.

Vidal, N., Peeters, M., Mulanga-Kabeya, C., Nzilambi, N., Robertson, D., Ilunga, W., Sema, H., Tshimanga, K., Bongo, B., Delaporte, E., 2000. Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. J. Virol.; 74: 10498-10507.

Wang, W-K., Chen, M-Y., Chuang, C-Y., Jeang, K-T., Huang, L-M., 2000. Molecular biology of human immunodeficiency virus type 1. J. Microbiol. Immunol. Infect.; 33: 131-140.

Wei, X., Ghosh, S. K., Taylor, M. E., Johnson, V. A., Emini, E. A., Deutsch, P., Lifson, J. D., Bonhoeffer, S., Nowak, M. A., Hahn, B. H., Shaw, G. M., 1995. Viral dynamics in human immunodeficiency virus type 1 infection. Nature; 373: 117-122.

WHO/UNAIDS 2007. Guidance on provider-initiated HIV testing and counselling in health facilities. Available: http://www.who.int/hiv/pub/vct/pitc2007/en/index.html. Accessed: 10 September 2012.

Wilkinson, E., Engelbrecht, S., 2009. Molecular characterization of non-subtype C and recombinant HIV-1 viruses from Cape Town, South Africa. Infection, Genetics and Evolution; 9: 840-846.

Williamson, C., Morris, L., Maughan, M. F., Ping, L. H., Dryga, S. A., Thomas, R., Reap, E. A., Cilliers, T., Van Harmelen, J., Pascual, A., Ramjee, G., Gray, G., Johnston, R., Karim, S. A., Swanstrom, R., 2003. Characterization and selection of HIV-1 subtype C isolates for use in vaccine development. AIDS Res. Human Retrovir.; 19: 133-144.

Williamson, C., Engelbrecht, S., Lambrick, M., Van Rensburg, E. J., Wood, R., Bredell, W., Williamson A-L., 1995. HIV-1 subtypes in different risk groups in South Africa Lancet 346: 782.

Worobey, M., Gemmel, M., Teuwen, D. E., Haselkorn, T., Kunstman, K., Bunce, M., Muyembe, J-J., Kabongo, J-M., Kalengayi, R. M., Van Marck, E., Gilbert, T. P., Wolinsky, S. M., 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature; 455: 661-664.

Yang, Z., 1994. Estimating the pattern of nucleotide substitution. J. Mol. Evol.; 39: 105-111.

Zhuang, J., Jetzt, A. E., Sun, G., Yu, H., Klarmann, G., Ron, Y., Preston, B. D., Dougherty, J. P., 2002. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. J. Virol.; 11273-11282.

Zur Megede, J., Engelbrecht, S., De Oliveira, T., Cassol, S., Scriba, T. J., Janse Van Rensburg, E., Barnett, S. W., 2002. Novel evolutionary analyses of full length HIV type 1 subtype C molecular clones from Cape Town, South Africa. AIDS Res. Human Retrovir.; 18: 1327-1332.

# APPENDIX

**Table 1.** LANL QC results of the *gag* sequences

| SeqName | Blast | Subtype | StopCodons | Frameshifts |
|---|---|---|---|---|
| 005a_gag | FJ198524\|ZA\|C\|96 | C | 0 | 1 |
| 0038A_gag | JN014194\|ZM\|C\|96 | C | 0 | 1 |
| 0039A_gag | FJ198843\|ZA\|C\|96 | C | 0 | 2 |
| 0040A_gag | GU017793\|UG\|C\|96 | C | 0 | 1 |
| 0042A_gag | HQ423975\|ZA\|C\|96 | B,C,F2,A2 | 0 | 1 |
| 0064A_gag | AF543954\|ZA\|C\|96 | C | 0 | 2 |
| 0066A_gag | AY255819\|IL\|C\|96 | C | 0 | 1 |
| 0073A_gag | FJ198849\|ZA\|C\|96 | C | 0 | 2 |
| 0081A_gag | FJ198691\|ZA\|C\|98 | C | 0 | 1 |
| 0085A_gag | FJ198491\|ZA\|C\|95 | C | 0 | 0 |
| 0092A_gag | FJ198550\|ZA\|C\|97 | C | 0 | 1 |
| 0097A_gag | FJ198629\|ZA\|C\|98 | C | 0 | 2 |
| 0098A_gag | AF543968\|ZA\|C\|96 | C | 0 | 2 |
| 0101A_gag | DQ093605\|ZA\|C\|95 | C | 0 | 1 |
| 0103A_gag | FJ199034\|ZA\|C\|97 | C | 0 | 1 |
| 0119A_gag | FJ606422\|ZM\|C\|94 | C | 0 | 1 |
| 0122A_gag | EU698192\|ZA\|C\|93 | C | 0 | 1 |
| 0123A_gag | FJ198849\|ZA\|C\|97 | C | 0 | 2 |
| 0130A_gag | JN167458\|ZA\|C\|98 | C | 0 | 1 |
| 0132A_gag | FJ606162\|ZM\|C\|97 | C | 0 | 0 |
| 0134A_gag | AF543951\|ZA\|C\|97 | C | 0 | 1 |
| 0135A_gag | DQ792993\|ZW\|C\|97 | C | 0 | 1 |
| 0143A_gag | HM593145\|ZA\|C\|95 | C | 0 | 1 |
| 0147A_gag | JN167458\|ZA\|C\|98 | C | 0 | 2 |
| 0152A_gag | EU698363\|ZA\|C\|97 | C | 1 | 2 |
| 0165A_gag | FJ198508\|ZA\|C\|98 | C | 1 | 2 |
| 0173A_gag | FJ198629\|ZA\|C\|98 | C | 0 | 2 |
| 0185A_gag | EU698192\|ZA\|C\|97 | C | 1 | 2 |
| 0189A_gag | AY838592\|ZA\|C\|97 | A1,A2,B,C | 0 | 2 |
| 0190A_gag | AF543954\|ZA\|C\|97 | C | 0 | 2 |
| 0192A_gag | FJ199034\|ZA\|C\|97 | C | 1 | 0 |
| 0193_gag | FJ198599\|ZA\|C\|97 | C | 0 | 2 |
| 0198A_gag | EF178402\|AU\|B\|89 | Cannot determine | 0 | 0 |
| 0199A_gag | FJ497890\|BW\|C\|97 | C | 0 | 2 |
| 0203A_gag | FJ198571\|ZA\|C\|98 | C,D,A2,01_AE,F2 | 0 | 2 |
| 0204A_gag | GU201724\|ZA\|C\|96 | C | 0 | 1 |
| 0206A_gag | AY713416\|SN\|C\|97 | C | 0 | 1 |
| 0207A_gag | AF543968\|ZA\|C\|96 | C | 0 | 1 |
| 0211A_gag | DQ093599\|ZA\|C\|96 | C | 0 | 1 |
| 0215A_gag | FJ198831\|ZA\|C\|98 | C | 0 | 2 |

**Table 2.** LANL QC results of the *pol (PR* and *RT)* sequences

| Name | Blast | RIP Subtype | Tree | Stop Codons | Frameshifts | Hypermutation |
|---|---|---|---|---|---|---|
| 0040_pol | FJ498439\|BW\|C\|95 | C | NJ Tree | 0 | 2 | Not Detected |
| 0042A_pol | FJ199945\|ZA\|C\|96 | C | NJ Tree | 0 | 5 | Not Detected |
| 0064A_pol | AY253317\|TZ\|C\|95 | C | NJ Tree | 0 | 4 | Not Detected |
| 0085_pol | GQ427130\|ZM\|C\|97 | C | NJ Tree | 0 | 3 | Not Detected |
| 0130A_pol | JN638131\|ZA\|C\|96 | C | NJ Tree | 3 | 7 | Not Detected |
| 0132A_pol | EF602226\|ZA\|C\|94 | C,F1 | NJ Tree | 0 | 2 | Not Detected |
| 0136A_pol | FJ199639\|ZA\|C\|96 | C | NJ Tree | 0 | 2 | Not Detected |
| 0143A_pol | DQ396391\|ZA\|C\|96 | C | NJ Tree | 0 | 1 | Not Detected |
| 0143A_44_pol | GQ433857\|ZM\|C\|95 | C | NJ Tree | 0 | 1 | Not Detected |
| 0173A_pol | HQ994716\|ZA\|C\|96 | C | NJ Tree | 0 | 5 | Not Detected |
| 0189A_pol | FJ199856\|ZA\|C\|96 | C | NJ Tree | 0 | 2 | Not Detected |
| 0192A_pol | AF443095\|BW\|C\|96 | C | NJ Tree | 0 | 2 | Not Detected |
| 0204A_pol | FJ199552\|ZA\|C\|90 | Cannot determine | NJ Tree | 0 | 0 | Not Detected |
| 0206A_pol | HQ994754\|ZA\|C\|94 | C | NJ Tree | 0 | 3 | Not Detected |
| 0207A_pol | EF602226\|ZA\|C\|94 | C | NJ Tree | 0 | 2 | Not Detected |

**Table 3.** LANL QC results of the *pol* IN sequences

| Name | Blast | RIP Subtype | Tree | Stop Codons | Frameshifts | Hypermutation |
|---|---|---|---|---|---|---|
| 0005A_IN | DQ369983\|ZA\|C\|96 | C | NJ Tree | 0 | 2 | Not Detected |
| 0040_IN | DQ375277\|SA\|C\|97 | C | NJ Tree | 0 | 2 | Not Detected |
| 0042A_IN | FJ480258\|ZA\|C\|95 | C | NJ Tree | 0 | 2 | Not Detected |
| 0064A_IN | DQ056415\|ZA\|C\|97 | C | NJ Tree | 0 | 1 | Not Detected |
| 0085A_IN | GQ999972\|ZA\|C\|96 | C | NJ Tree | 0 | 3 | Not Detected |
| 0130A_IN | AF110974\|BW\|C\|95 | C | NJ Tree | 0 | 1 | Possible |
| 0136A_IN | FJ199590\|ZA\|C\|97 | C | NJ Tree | 0 | 1 | Not Detected |
| 0143A_1_IN | AY463227\|ZA\|C\|97 | C | NJ Tree | 0 | 1 | Not Detected |
| 0143A_2_IN | DQ275659\|ZA\|C\|96 | C | NJ Tree | 0 | 3 | Not Detected |
| 0173A_IN | DQ275659\|ZA\|C\|96 | C | NJ Tree | 2 | 6 | Not Detected |
| 0189A_IN | FJ199670\|ZA\|C\|97 | C | NJ Tree | 0 | 1 | Not Detected |
| 0192A_IN | FJ480257\|ZA\|C\|95 | C,01_AE | NJ Tree | 1 | 3 | Not Detected |
| 0206A_IN | FJ480253\|ZA\|C\|96 | C | NJ Tree | 0 | 1 | Not Detected |
| 0207A_IN | FJ199553\|ZA\|C\|97 | C | NJ Tree | 0 | 2 | Possible |
| 0066A_IN | DQ369977\|ZA\|C\|97 | C | NJ Tree | 0 | 1 | Not Detected |
| 0081A_IN | DQ351237\|ZA\|C\|95 | C | NJ Tree | 0 | 0 | Not Detected |
| 0098A_IN | AF411965\|ZA\|A2C\|95 | C | NJ Tree | 0 | 1 | Not Detected |
| 0101A_IN | GQ872480\|ZA\|C\|96 | C | NJ Tree | 0 | 1 | Not Detected |
| 0185A_IN | DQ275659\|ZA\|C\|97 | C | NJ Tree | 0 | 1 | Not Detected |
| 0192A_IN22 | GQ872500\|ZA\|C\|97 | C | NJ Tree | 0 | 2 | Not Detected |
| 0193A_IN | HM569277\|ZA\|C\|96 | C | NJ Tree | 0 | 4 | Not Detected |
| 0198A_IN | HM026822\|BR\|C\|96 | C | NJ Tree | 0 | 0 | Not Detected |

**Table 4.** LANL QC results of the *env* sequences

| Name | Blast | RIP Subtype | Tree | Stop Codons | Frameshifts | Hypermutation |
|---|---|---|---|---|---|---|
| 0005_gp | HM638718\|MW\|C\|93 | C | NJ Tree | 0 | 2 | Not Detected |
| 0018A_gp | DQ866353\|ZA\|C\|92 | C | NJ Tree | 0 | 2 | Not Detected |
| 0022A_gp | AF391247\|ZA\|C\|99 | C | NJ Tree | 0 | 2 | Not Detected |
| 0039A_gp | DQ275660\|ZA\|C\|93 | C | NJ Tree | 0 | 2 | Not Detected |
| 0040_gp | HQ595761\|ZA\|C\|93 | C | NJ Tree | 0 | 1 | Not Detected |
| 0042A_gp | AY522732\|ZA\|C\|94 | C | NJ Tree | 0 | 2 | Not Detected |
| 0043A_gp | JQ609961\|US\|B\|90 | B | NJ Tree | 0 | 1 | Not Detected |
| 0055A_gp | AF391247\|ZA\|C\|99 | C | NJ Tree | 0 | 2 | Not Detected |
| 0064A_gp | AF286234\|TZ\|C\|94 | C | NJ Tree | 0 | 1 | Not Detected |
| 0066A_gp | GU329049\|ZM\|C\|94 | C | NJ Tree | 0 | 2 | Not Detected |
| 0073_gp | AY118165\|ZA\|C\|95 | C | NJ Tree | 0 | 2 | Not Detected |
| 0073A_gp | AY118165\|ZA\|C\|95 | C | NJ Tree | 0 | 1 | Not Detected |
| 0081A_gp | AY734558\|TZ\|C\|94 | C | NJ Tree | 0 | 1 | Not Detected |
| 0092A_gp | AF443084\|BW\|C\|94 | C | NJ Tree | 0 | 2 | Not Detected |
| 0097A_gp | DQ056412\|ZA\|C\|94 | C | NJ Tree | 0 | 2 | Not Detected |
| 0098A_gp | AY772695\|ZA\|C\|95 | C | NJ Tree | 0 | 2 | Not Detected |
| 0101A_gp | HM638816\|MW\|C\|92 | C | NJ Tree | 0 | 2 | Not Detected |
| 0103A_gp | DQ369988\|ZA\|C\|93 | C | NJ Tree | 0 | 1 | Not Detected |
| 0122_gp | DQ367245\|IN\|C\|94 | C | NJ Tree | 0 | 2 | Not Detected |
| 0123A_gp | DQ866337\|ZA\|C\|89 | C | NJ Tree | 0 | 2 | Not Detected |
| 0130A_gp | FJ846628\|ZA\|C\|93 | C | NJ Tree | 0 | 2 | Not Detected |
| 0132A_gp | AY713413\|MW\|C\|93 | C | NJ Tree | 0 | 2 | Not Detected |
| 0134A_gp | DQ056409\|ZA\|C\|96 | C | NJ Tree | 0 | 2 | Not Detected |
| 0143A | JQ609961\|US\|B\|92 | Cannot determine | NJ Tree | 0 | 0 | Not Detected |
| 0147_gp | AY901974\|ZA\|C\|94 | C | NJ Tree | 0 | 2 | Not Detected |
| 0152A_gp | U08455\|MW\|C\|93 | C | NJ Tree | 0 | 2 | Not Detected |
| 0155A_gp | DQ351229\|ZA\|C\|94 | C | NJ Tree | 0 | 1 | Not Detected |
| 0165A_gp | HM638824\|MW\|C\|95 | C | NJ Tree | 0 | 2 | Not Detected |
| | AY522736\|ZA\|C\|96 | C | NJ Tree | 0 | 2 | Not Detected |
| 0173A_gp | | | | | | |
| 0185A_gp | HM036835\|ZM\|C\|92 | C | NJ Tree | 0 | 2 | Not Detected |
| 0189_gp | DQ866353\|ZA\|C\|92 | C | NJ Tree | 0 | 2 | Not Detected |
| 0190A_gp | DQ866335\|ZA\|C\|95 | C | NJ Tree | 0 | 2 | Not Detected |
| 0192A_gp | DQ866390\|ZA\|C\|94 | C | NJ Tree | 0 | 2 | Not Detected |
| 0193_gp | HM215437\|ZA\|C\|96 | C | NJ Tree | 0 | 1 | Not Detected |
| 0193A_gp | DQ396373\|ZA\|C\|96 | C | NJ Tree | 0 | 2 | Not Detected |
| 0198A_gp_1 | HM638665\|MW\|C\|93 | C | NJ Tree | 0 | 1 | Not Detected |
| 0198A_gp_2 | HM638816\|MW\|C\|93 | C | NJ Tree | 0 | 2 | Not Detected |
| 0199A_gp_2 | AF443099\|BW\|C\|92 | C | NJ Tree | 0 | 2 | Not Detected |
| 0199A_gp_1 | DQ275663\|ZA\|C\|95 | C | NJ Tree | 0 | 2 | Not Detected |
| 0203A_gp | JQ715376\|MW\|C\|91 | C | NJ Tree | 0 | 2 | Not Detected |
| 0204A_gp | HQ707871\|ZW\|C\|92 | C | NJ Tree | 0 | 2 | Not Detected |
| 0206A_gp | AF219269\|FI\|C\|94 | C | NJ Tree | 0 | 1 | Not Detected |
| 0207_gp | FJ443426\|ZA\|C\|94 | C | NJ Tree | 0 | 2 | Not Detected |
| 0211_gp | GU216844\|ZA\|C\|93 | C | NJ Tree | 0 | 2 | Not Detected |

**Table 5.** REGA v 3.0 results of the *gag* sequences

| Name | Length bp | Subtype assignment | Support | Begin | End |
|---|---|---|---|---|---|
| 005a_gag | 458 | HIV-1 Subtype C | 98 | 1248 | 1706 |
| 0038A_gag | 459 | HIV-1 Subtype C | 100 | 1252 | 1711 |
| 0039A_gag | 454 | HIV-1 Subtype C | 95 | 1252 | 1706 |
| 0040A_gag | 458 | HIV-1 Subtype C | 96 | 1248 | 1706 |
| 0042A_gag | 846 | Recombinant of C, F1 | | 861 | 1707 |
| 0064A_gag | 455 | HIV-1 Subtype C | 98 | 1253 | 1708 |
| 0066A_gag | 456 | HIV-1 Subtype C | 98 | 1251 | 1707 |
| 0073A_gag | 454 | HIV-1 Subtype C | 99 | 1252 | 1706 |
| 0081A_gag | 452 | HIV-1 Subtype C | 100 | 1249 | 1701 |
| 0085A_gag | 481 | HIV-1 Subtype C | 100 | 1248 | 1729 |
| 0092A_gag | 455 | HIV-1 Subtype C | 100 | 1251 | 1706 |
| 0097A_gag | 459 | HIV-1 Subtype C | 99 | 1247 | 1706 |
| 0098A_gag | 457 | HIV-1 Subtype C | 100 | 1249 | 1706 |
| 0101A_gag | 458 | HIV-1 Subtype C | 96 | 1248 | 1706 |
| 0103A_gag | 456 | HIV-1 Subtype C | 98 | 1250 | 1706 |
| 0119A_gag | 458 | Check the report | | 1248 | 1706 |
| 0122A_gag | 425 | HIV-1 Subtype C | 98 | 1248 | 1673 |
| 0123A_gag | 457 | HIV-1 Subtype C | 99 | 1249 | 1706 |
| 0130A_gag | 459 | HIV-1 Subtype C | 98 | 1248 | 1707 |
| 0132A_gag | 484 | HIV-1 Subtype C | 99 | 1248 | 1732 |
| 0134A_gag | 458 | HIV-1 Subtype C | 99 | 1248 | 1706 |
| 0135A_gag | 458 | HIV-1 Subtype C | 100 | 1251 | 1709 |
| 0143A_gag | 482 | HIV-1 Subtype C | 93 | 1248 | 1730 |
| 0147A_gag | 467 | HIV-1 Subtype C | 100 | 1240 | 1707 |
| 0152A_gag | 462 | HIV-1 Subtype C | 98 | 1249 | 1711 |
| 0165A_gag | 459 | HIV-1 Subtype C | 96 | 1248 | 1706 |
| 0173A_gag | 460 | HIV-1 Subtype C | 100 | 1246 | 1706 |
| 0185A_gag | 457 | HIV-1 Subtype C | 100 | 1251 | 1706 |
| 0189A_gag | 882 | Subtype C, potential recombinant | 94 | 826 | 1708 |
| 0190A_gag | 455 | HIV-1 Subtype C | 98 | 1252 | 1707 |
| 0192A_gag | 456 | HIV-1 Subtype C | 98 | 1251 | 1706 |
| 0193_gag | 456 | HIV-1 Subtype C | 98 | 1250 | 1706 |
| 0198A_gag | 458 | Check the report | | 1250 | 1708 |
| 0199A_gag | 459 | HIV-1 Subtype C | 100 | 1249 | 1708 |
| 0203A_gag | 863 | HIV-1 Subtype C | 94 | 843 | 1706 |
| 0204A_gag | 464 | HIV-1 Subtype C | 99 | 1245 | 1709 |
| 0206A_gag | 424 | HIV-1 Subtype C | 91 | 1284 | 1708 |
| 0207A_gag | 458 | HIV-1 Subtype C | 97 | 1248 | 1706 |
| 0211A_gag | 458 | HIV-1 Subtype C | 100 | 1248 | 1706 |
| 0215A_gag | 456 | HIV-1 Subtype C | 100 | 1250 | 1706 |

**Table 6.** REGA v 3.0 results of the *pol (PR* and *RT)* sequences

| Name | Length bp | Subtype assignment | Support | Begin | End |
|---|---|---|---|---|---|
| 0040_pol | 1215 | HIV-1 Subtype C | 100 | 2108 | 3335 |
| 0042A_pol | 1215 | HIV-1 Subtype C | 100 | 2111 | 3335 |
| 0064A_pol | 1209 | HIV-1 Subtype C | 100 | 2132 | 3339 |
| 0085_pol | 1204 | HIV-1 Subtype C | 100 | 2131 | 3340 |
| 0130A_pol | 1201 | HIV-1 Subtype C | 100 | 2132 | 3339 |
| 0132A_pol | 1319 | HIV-1 Subtype C | 100 | 2132 | 3460 |
| 0136A_pol | 1189 | HIV-1 Subtype C | 100 | 2131 | 3335 |
| 0143A_pol | 1181 | HIV-1 Subtype C | 100 | 2142 | 3335 |
| 0143A_44_pol | 1216 | HIV-1 Subtype C | 100 | 2112 | 3340 |
| 0173A_pol | 1206 | HIV-1 Subtype C | 100 | 2124 | 3335 |
| 0189A_pol | 711 | HIV-1 Subtype C | 100 | 2132 | 2855 |
| 0192A_pol | 625 | HIV-1 Subtype C | 100 | 2134 | 2771 |
| 0204A_pol | 351 | Check the report | | 2177 | 2534 |
| 0206A_pol | 1202 | HIV-1 Subtype C | 100 | 2113 | 3330 |
| 0207A_pol | 1196 | HIV-1 Subtype C | 100 | 2133 | 3341 |

**Table 7.** REGA v 3.0 results of the pol *IN gag* sequences

| Name | Length bp | Subtype assignment | Support | Begin | End |
|---|---|---|---|---|---|
| 0005A_IN | 878 | HIV-1 Subtype C | 100 | 4206 | 5084 |
| 0040_IN | 907 | HIV-1 Subtype C | 100 | 4207 | 5114 |
| 0042A_IN | 910 | HIV-1 Subtype C | 100 | 4203 | 5113 |
| 0064A_IN | 908 | HIV-1 Subtype C | 100 | 4203 | 5111 |
| 0085A_IN | 903 | HIV-1 Subtype C | 100 | 4201 | 5104 |
| 0130A_IN | 915 | HIV-1 Subtype C | 100 | 4200 | 5115 |
| 0136A_IN | 983 | HIV-1 Subtype C | 100 | 4207 | 5190 |
| 0143A_1_IN | 890 | HIV-1 Subtype C | 100 | 4213 | 5103 |
| 0143A_2_IN | 880 | HIV-1 Subtype C | 100 | 4203 | 5083 |
| 0173A_IN | 884 | HIV-1 Subtype C | 100 | 4213 | 5097 |
| 0189A_IN | 900 | HIV-1 Subtype C | 100 | 4208 | 5108 |
| 0192A_IN | 895 | HIV-1 Subtype C | 100 | 4206 | 5102 |
| 0206A_IN | 892 | HIV-1 Subtype C | 100 | 4208 | 5100 |
| 0207A_IN | 906 | HIV-1 Subtype C | 100 | 4207 | 5113 |
| 0066A_IN | 871 | HIV-1 Subtype C | 100 | 4211 | 5082 |
| 0081A_IN | 868 | HIV-1 Subtype C | 100 | 4229 | 5097 |
| 0098A_IN | 849 | HIV-1 Subtype C | 100 | 4235 | 5084 |
| 0101A_IN | 860 | HIV-1 Subtype C | 100 | 4247 | 5107 |
| 0185A_IN | 875 | HIV-1 Subtype C | 100 | 4223 | 5098 |
| 0192A_IN | 866 | HIV-1 Subtype C | 100 | 4229 | 5095 |
| 0193A_IN | 780 | HIV-1 Subtype C | 100 | 4226 | 5006 |
| 0198A_IN | 817 | HIV-1 Subtype C | 100 | 4226 | 5043 |

**Table 8.** REGA v 3.0 results of the *env* sequences

| Name | Length bp | Subtype assignment | Support | Begin | End |
|---|---|---|---|---|---|
| 0005_gp | 408 | HIV-1 Subtype C | 99 | 7875 | 8283 |
| 0018A_gp | 399 | HIV-1 Subtype C | 100 | 7881 | 8280 |
| 0022A_gp | 424 | HIV-1 Subtype C | 100 | 7872 | 8296 |
| 0039A_gp | 421 | HIV-1 Subtype C | 100 | 7875 | 8296 |
| 0040_gp | 404 | HIV-1 Subtype C | 100 | 7878 | 8282 |
| 0042A_gp | 408 | HIV-1 Subtype C | 100 | 7875 | 8283 |
| 0043A_gp | 400 | HIV-1 Subtype B | 92 | 7895 | 8284 |
| 0055A_gp | 411 | HIV-1 Subtype C | 99 | 7875 | 8286 |
| 0064A_gp | 373 | HIV-1 Subtype C | 100 | 7914 | 8287 |
| 0066A_gp | 409 | HIV-1 Subtype C | 99 | 7878 | 8287 |
| 0073_gp | 422 | HIV-1 Subtype C | 100 | 7872 | 8294 |
| 0073A_gp | 410 | HIV-1 Subtype C | 100 | 7878 | 8288 |
| 0081A_gp | 410 | HIV-1 Subtype C | 100 | 7875 | 8285 |
| 0092A_gp | 409 | HIV-1 Subtype C | 100 | 7875 | 8284 |
| 0097A_gp | 409 | HIV-1 Subtype C | 99 | 7875 | 8284 |
| 0098A_gp | 409 | HIV-1 Subtype C | 100 | 7875 | 8284 |
| 0101A_gp | 406 | HIV-1 Subtype C | 100 | 7878 | 8284 |
| 0103A_gp | 410 | HIV-1 Subtype C | 99 | 7875 | 8285 |
| 0122_gp | 439 | HIV-1 Subtype C | 99 | 7863 | 8302 |
| 0123A_gp | 359 | HIV-1 Subtype C | 100 | 7925 | 8286 |
| 0130A_gp | 409 | HIV-1 Subtype C | 100 | 7875 | 8284 |
| 0132A_gp | 405 | HIV-1 Subtype C | 100 | 7878 | 8283 |
| 0134A_gp | 408 | HIV-1 Subtype C | 100 | 7875 | 8283 |
| 0143A | 386 | Check the report | | 7867 | 8241 |
| 0147_gp | 409 | HIV-1 Subtype C | 100 | 7875 | 8284 |
| 0152A_gp | 409 | HIV-1 Subtype C | 100 | 7875 | 8284 |
| 0155A_gp | 408 | HIV-1 Subtype C | 100 | 7875 | 8283 |
| 0165A_gp | 408 | HIV-1 Subtype C | 100 | 7875 | 8283 |
| 0173A_gp | 400 | HIV-1 Subtype C | 100 | 7875 | 8275 |
| 0185A_gp | 413 | HIV-1 Subtype C | 100 | 7875 | 8288 |
| 0189_gp | 370 | HIV-1 Subtype C | 96 | 7914 | 8284 |
| 0190A_gp | 406 | HIV-1 Subtype C | 100 | 7878 | 8284 |
| 0192A_gp | 408 | HIV-1 Subtype C | 100 | 7875 | 8283 |
| 0193_gp | 422 | HIV-1 Subtype C | 100 | 7862 | 8284 |
| 0193A_gp | 408 | HIV-1 Subtype C | 97 | 7875 | 8283 |
| 0198A_gp_1 | 405 | HIV-1 Subtype C | 100 | 7878 | 8283 |
| 0198A_gp_2 | 393 | HIV-1 Subtype C | 100 | 7890 | 8283 |
| 0199A_gp_2 | 409 | HIV-1 Subtype C | 99 | 7875 | 8284 |
| 0199A_gp_1 | 406 | HIV-1 Subtype C | 99 | 7878 | 8284 |
| 0203A_gp | 405 | HIV-1 Subtype C | 98 | 7878 | 8283 |
| 0204A_gp | 412 | HIV-1 Subtype C | 100 | 7872 | 8284 |
| 0206A_gp | 408 | HIV-1 Subtype C | 99 | 7875 | 8283 |
| 0207_gp | 406 | HIV-1 Subtype C | 99 | 7878 | 8284 |
| 0211_gp | 409 | HIV-1 Subtype C | 100 | 7875 | 8284 |

**Table 9.** SQUEAL results of the *pol (PR* and *RT)* sequences

| Name | Subtype | Expanded subtype | Confidence in Assignment | Support for recombination | Support for intra-subtype recombination | Breakpoints |
|---|---|---|---|---|---|---|
| 0040_pol | C | C | 0.764177 | 0.235823 | 0.235816 | |
| 0042A_pol | C | C intra-subtype recombinant (2 breakpoints) | 0.556369 | 0.996056 | 0.995983 | 819(817-821); 944(941-947) |
| 0064A_pol | C | C | 0.927817 | 0.0721828 | 0.0721729 | |
| 0085_pol | C | C | 0.999544 | 0.000456014 | 0.000456014 | |
| 0130A_pol | C | C | 0.835575 | 0.164425 | 0.0509365 | |
| 0132A_pol | C,CPZ | C,C,C,CPZ inter-subtype recombinant | 0.632129 | 1 | 0.000211618 | 101(1-492); 499(498-500); 1094(1093-1095) |
| 0136A_pol | C | C | 0.999801 | 0.000199417 | 0.000199417 | |
| 0143A_pol | C | C intra-subtype recombinant (4 breakpoints) | 0.50413 | 0.999791 | 0.997879 | 101(50-152); 643(642-644); 814(813-815); 956(955-957) |
| 0143A_44_pol | C | C | 0.99996 | 4.05E-05 | 4.05E-05 | |
| 0173A_pol | C | C intra-subtype recombinant (1 breakpoints) | 0.664622 | 0.79679 | 0.796765 | 316(232-400) |
| 0189A_pol | C | C | 0.695241 | 0.304759 | 0.304759 | |
| 0192A_pol | C | C | 0.940919 | 0.0590808 | 0.0590808 | |
| 0206A_pol | C | C | 0.673925 | 0.326075 | 0.326074 | |
| 0207A_pol | C | C intra-subtype recombinant (2 breakpoints) | 0.786224 | 0.790867 | 0.790866 | 110(109-111); 284(271-297) |

**Table 10.** SQUEAL results of the *pol (IN)* sequences

| Name | Expanded subtype | Confidence in Assignment | Support for recombination | Support for intra-subtype recombination | Breakpoints |
|---|---|---|---|---|---|
| 0005A_IN | C | 0.992215 | 0.00778526 | 0.00778526 | |
| 0040_IN | C intra-subtype recombinant (1 breakpoints) | 0.724519 | 0.758161 | 0.758152 | 555(466-644) |
| 0042A_IN | C | 0.994517 | 0.00548278 | 0.00548278 | |
| 0064A_IN | C | 0.609436 | 0.390564 | 0.390389 | |
| 0085A_IN | C | 0.999718 | 0.000281964 | 0.000281964 | |
| 0130A_IN | C | 0.646731 | 0.353268 | 0.353268 | |
| 0136A_IN | C | 0.799613 | 0.200387 | 0.200384 | |
| 0143A_1_IN | C intra-subtype recombinant (2 breakpoints) | 0.807617 | 0.967812 | 0.967722 | 101(97-105); 249(237-261) |
| 0143A_2_IN | C | 0.999955 | 4.53E-05 | 4.53E-05 | |
| 0173A_IN | C intra-subtype recombinant (2 breakpoints) | 0.662959 | 0.679639 | 0.679073 | 267(258-276); 564(562-566) |
| 0189A_IN | C | 0.999969 | 3.11E-05 | 2.46E-05 | |
| 0192A_IN | C intra-subtype recombinant (3 breakpoints) | 0.557236 | 0.999569 | 0.981578 | 265(262-268); 409(408-410); 774(773-775) |
| 0206A_IN | C intra-subtype recombinant (1 breakpoints) | 0.614465 | 0.675081 | 0.675066 | 784(778-790) |
| 0207A_IN | C | 0.998927 | 0.00107285 | 0.00107272 | |
| 0066A_IN | C | 0.970445 | 0.0295545 | 0.0295545 | |
| 0081A_IN | C intra-subtype recombinant (2 breakpoints) | 0.533426 | 0.599499 | 0.59949 | 408(402-414); 637(622-652) |
| 0098A_IN | C intra-subtype recombinant (3 breakpoints) | 0.86141 | 0.999989 | 0.999773 | 172(171-173) 383(382-384); 736(714-758) |
| 0101A_IN | C | 0.997377 | 0.00262261 | 0.00262261 | |
| 0185A_IN | C | 0.999237 | 0.000762537 | 0.000762537 | |
| 0192A_IN | C | 0.861912 | 0.138088 | 0.138088 | |
| 0193A_IN | C | 0.997888 | 0.00211194 | 0.00211194 | |
| 0198A_IN | C | 0.991374 | 0.00862639 | 0.00847651 | |

**Table 11.** SDRM results of the *pol* sequences

| **Number of input sequences** | | | |
|---|---|---|---|
| | **No. submitted** | **No. filtered (QA)*** | **No. analyzed for SDRMs** |
| Sequences | 15 | | |
| Sequences containing either RT or PR | 15 | 1 | 14 |
| Sequences containing RT | 13 | 1 | 12 |
| Sequences containing PR | 15 | 1 | 14 |
| Sequences containing both RT and PR | 13 | 1 | 12 |
| **Proportion of sequences with SDRMs** | | | |
| **Resistance category** | **No. analyzed** | **No. containing SDRM** | **%** |
| Sequences with any SDRM | 14 | 2 | 14.3 |
| PR Sequences with any PI SDRM | 14 | 0 | 0 |
| RT Sequences with any NRTI SDRM | 12 | 1 | 8.3 |
| RT Sequences with any NNRTI SDRM | 12 | 1 | 8.3 |
| RT Sequences with any NRTI + any NNRTI SDRM | 12 | 0 | 0 |
| PRRT Sequences with any NRTI + any NNRTI + PI SDRM | 12 | 0 | 0 |
| **Sequences with SDRMs** | | | |
| **SequenceID** | **NRTI SDRMs** | **NNRTI SDRMs** | **PI SDRMs** |
| 0130A_pol | T215S | None | None |
| 0143A_44_pol | None | K103N | None |
| | | | |